

W P 5 R E H (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

Mrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Mar 19 09:53:17 1996; MasPar time 7.39 Seconds
555.014 Million cell updates/sec

Tabular output not generated.
Title: >US-08-404-832-2
Description: (1-567) from US08404832.pep
Perfect Score: 4162
Sequence: 1 MESSKMDSPGALQTNPLK.....IKDDTIFIKVYDTSDDPP 567

Scoring table: PAM 150
Gap 11
Searched: 62355 seqs, 7230759 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq20
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12

Statistics: Mean 36.298; Variance 168.378; scale 0.216

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	132	3.2	527	7 R38099	Schistosoma mansoni	8.41e-02
2	132	3.2	527	3 R4163	Product of clone Irv-	1.41e-02
3	128	3.1	235	1 P94260	41KD protein of T. co	1.63e-01
4	128	3.1	320	1 P94366	41KD protein of T. co	1.63e-01
5	130	3.1	885	12 R6930	AMML chromosome inv(1	1.17e-01
6	125	3.0	1284	1 P8187	Sequence encoded by a	2.66e-01
7	118	2.8	252	1 P94369	Fusion protein congt.	8.27e-01
8	116	2.8	334	10 R51227	Membrane antigen pept	1.14e+00
9	117	2.8	337	11 R63681	Merosin amino acids 1	9.71e-01

10	117	2.8	385	1 P90994	B. subtilis sacU1 po	9.71e-01
11	118	2.8	476	8 R43563	Hyaluronan receptor.	8.27e-01
12	114	2.7	462	1 R05766	Portion of peptide an	1.57e+00
13	111	2.7	769	1 P90954	Yeast topoisomerase I	2.52e+00
14	111	2.7	870	6 R30729	p100 protein from hum	2.52e+00
15	107	2.6	475	3 R15148	Ro/SSA autoantigen.	4.70e+00
16	110	2.6	816	12 R66931	AMML chromosome inv(1	2.94e+00
17	108	2.6	3685	1 P90290	Human Duchenne muscul	4.02e+00
18	103	2.5	200	9 R47245	Fragment of 101 kD pr	8.71e+00
19	103	2.5	318	5 R26943	P. falciparum LSA N-te	8.71e+00
20	104	2.5	674	6 R34130	Truncated PVX replica	7.47e+00
21	104	2.5	707	6 R34131	Truncated PVX replica	7.47e+00
22	104	2.5	1456	6 R34129	PVX replicase.	7.47e+00
23	104	2.5	1618	5 R27205	Human nestin.	7.47e+00
24	104	2.5	1618	11 R60127	Human nestin protein	7.47e+00
25	98	2.4	183	7 R39482	Human apolipoprotein	1.86e+01
26	101	2.4	316	5 R26941	P. falciparum LSA-NR	1.86e+01
27	98	2.4	333	7 R39480	Human apolipoprotein	1.86e+01
28	98	2.4	337	7 R39492	Human apolipoprotein	1.86e+01
29	98	2.4	342	7 R39487	Human apolipoprotein	1.86e+01
30	98	2.4	342	7 R39496	Human apolipoprotein	1.86e+01
31	98	2.4	342	7 R39491	Human apolipoprotein	1.86e+01
32	98	2.4	346	7 R39493	Human apolipoprotein	1.86e+01
33	99	2.4	361	9 R31052	Sequence encoded by t	1.60e+01
34	98	2.4	363	7 R39479	Human apolipoprotein	1.86e+01
35	98	2.4	373	7 R39486	Human apolipoprotein	1.86e+01
36	98	2.4	377	7 R39443	Human apolipoprotein	1.86e+01
37	98	2.4	377	7 R45244	Human apolipoprotein	1.86e+01
38	98	2.4	377	7 R39502	Human apolipoprotein	1.86e+01
39	98	2.4	377	7 R39501	Human apolipoprotein	1.86e+01
40	98	2.4	377	7 R45243	Human apolipoprotein	1.86e+01
41	98	2.4	377	7 R45242	Human apolipoprotein	1.86e+01
42	101	2.4	493	5 R26944	P. falciparum LSA gene	1.18e+01
43	99	2.4	576	12 R66929	AMML chromosome inv(1	1.60e+01
44	99	2.4	955	11 R57365	K39 polypeptide of le	1.60e+01
45	99	2.4	2101	9 R47173	Sequence of the inner	1.60e+01

ALIGNMENTS

RESULT 1
ID R38099 standard; Protein; 527 AA.
AC R38099;
DT 26-OCT-1993 (first entry)
DE Schistosoma mansoni epitope;
KW Antibody; protein epitope; glycanic epitope; surface;
KW schistosoma; myosin; C. elegans.
OS Schistosoma mansoni.
PN US5219566-A.
PD 15-JUN-1993.
PF 30-SEP-1988; 252075.
PR 30-SEP-1988; US-252075.
PR 09-JUL-1991; US-725804.
PA (U100) UNIV JOHNS HOPKINS.
PI Strand M;
PI WPI; 93-205255/25.
DR N-PEDB; Q43521.
DR New cDNA encoding polypeptide of Schistosoma mansoni - includes
PT protein epitope(s) present on Schistosoma surface, used in
PT vaccines against bilharziasis

PS Claim 1; Fig 1B; 11pp; English.
CC The polypeptide binds to antibodies which recognise protein epitopes,
CC but not glycanic epitopes, expressed on the surface of live
CC schistosoma of S. mansoni. The protein epitopes are common to a
CC 200 kD and a 38 kD glycoprotein of S. mansoni. These glycoproteins
CC are immunologically cross-reactive with myosin heavy chains from other
CC species. However, anti-myosin antibodies directed against myosin mols.
CC of other species are not cross-reactive with the surface epitopes of
CC S. mansoni.
SQ Sequence 527 AA;

Query Match 3.2%, Score 132; DB 7; Length 527;
Best Local Similarity 24.4%; Pred. No. 8.41e-02;
Matches 29; Conservative 33; Mismatches 51; Indels 6; Gaps 4;

Db 255 dneirtqsemagdemgklnkdknleeeenkrtrgealgaedkvnhlklakleat1 314
QY 270 EKVSLLONEVEKNKKSISQILHNOICFELIEIRKEMLRNNSKILHQRVIDSOAKL 329
Db 315 demeenalareqkirkdvskirkleg-dlkatq-etvd-dlervkrd--leeqlrke 367
QY 330 KEIDKEIRPFRRNWEADSMKSSVESIQNRVTELESVDKSAGVANTLTLESQSRD 388

RESULT 2
ID R14183 standard; Protein; 527 AA.
AC R14183;
DT 03-JAN-1992 (first entry)
DE Product of clone IRV-5.
KW Epitope; bilharziasis; antibodies; schistosomula; vaccine;
KW immunogen; anti-IRV; fusion.
OS Schistosoma mansoni.
FH Key Location/Qualifiers
FT Domain 121..572
FT /label= alpha helix
FT Modified-site 160
FT /label= N-glycosite
FT Modified-site 218
FT /label= N-glycosite
PN US5051254-A.
PD 24-SEP-1991.
PF 30-SEP-1988; 252075.
PR 30-SEP-1988; US-252075.
PA (UYJO) JOHNS HOPKINS UNIV.
PI Strand M;
PI WPI; 91-303003/41.
DR N-PSDB; Q13982.
PT Polypeptide(s) specific for epitope(s) of Schistosoma mansoni -
PT used as vaccines against bilharziasis.
PS Claim 5; Fig 1B; 10pp; English.
CC The clone was obt. by screening an adult worm cDNA library in
CC lambda-gt10 with anti-IRV sera raised against a subset of adult worm
CC antigens. The sequence is thought to at least partially encode a
CC schistosome myosin. It has 48% identity with the myosin heavy
CC chain of C. elegans. The DNA can be used to produce a fusion protein
CC with beta galactosidase sequences. The product can be used as an
CC immunogen to produce vaccines against bilharziasis.
SQ Sequence 527 AA;

Query Match 3.2%, Score 132; DB 3; Length 527;

Best Local Similarity 24.4%; Pred. No. 8.41e-02;
Matches 29; Conservative 33; Mismatches 51; Indels 6; Gaps 4;

Db 255 dneirtqsemagdemgklnkdknleeeenkrtrgealgaedkvnhlklakleat1 314
QY 270 EKVSLLONEVEKNKKSISQILHNOICFELIEIRKEMLRNNSKILHQRVIDSOAKL 329
Db 315 demeenalareqkirkdvskirkleg-dlkatq-etvd-dlervkrd--leeqlrke 367
QY 330 KEIDKEIRPFRRNWEADSMKSSVESIQNRVTELESVDKSAGVANTLTLESQSRD 388

RESULT 3
ID P94260 standard; protein; 235 AA.
AC P94260;
DT 13-JUN-1990 (first entry)
DE 41kD protein of T. colubriformis.
KW Parasitic nematode; 41 kD protein.
OS Trichostrongylus colubriformis.
PN WO8900163-A.
PD 12-JAN-1989.
PF 06-JUL-1988; A00239.
PR 07-JUL-1987; AU-002940.
PA (BIOT-) Biotech Austr Pt, (CSIR) Commonwealth Sci and Ind Res Orgn.
PI Cobon GS, Austen RA, O'Donnell JU, Frenkel MJ, Kennedy WK, Savin KW,
PI Wagland BM;
DR N-PSDB; N91210.
DR WPI; 89-039628/05.
PT Protein derived from parasitic nematode species - used to provide
PT protective immunity against nematode parasites of man and animal.s
PS Disclosure; 57pp; English.
CC Recombinant organisms contg. DNA encoding the protein can be used to
CC produce polypeptide which is capable of protecting guinea pigs
CC against parasitism from T. colubriformis.
CC See also P94366-P94369.
SQ Sequence 235 AA;

Query Match 3.1%, Score 128; DB 1; Length 235;
Best Local Similarity 17.1%; Pred. No. 1.63e-01;
Matches 18; Conservative 35; Mismatches 49; Indels 3; Gaps 2;

Db 7 mkiekdnaldradaaeekvrgiteklerveeirdtqkmmgtendldkagelaaatsq 66
QY 299 IEIERKEMLRN--ESKILHQRVIDSOAKELKEIRPFRRNWEA--DSMKSSVES 355
Db 67 leekkyggaeeaaalnrmclleeeleeraerlkiateklee 111
QY 356 LONRVTELESVDKSAGVANTLTLESQSRDQMLSVHDIRLAD 400

RESULT 4
ID P94366 standard; protein; 320 AA.
AC P94366;
DT 13-JUN-1990 (first entry)
DE 41kD protein of T. colubriformis.
KW Parasitic nematode; 41 kD protein.
OS Trichostrongylus colubriformis.
PN WO8900163-A.
PD 12-JAN-1989.
PF 06-JUL-1988; A00239.

PR 07-JUL-1987; AU-002940.
PA (BIOT-) Biotech Aust Pt. (CSIR) Commonwealth Sci and Ind Res Orgn.
PI Cobon GS, Austen RA, O'Donnell LJ, Frenkel MJ, Kennedy WPK, Savin KM,
PI Magland BM;
DR WPI; 89-039628/05.
DR N-PSDB; N91211.
PR Protein derived from parasitic nematode species - used to provide
PT protective immunity against nematode parasites of man and animals.
PS Claim 4; page 39; 57pp; English.
CC Recombinant organisms conty. DNA encoding the protein can be used to
CC produce polypeptide which is capable of protecting guinea pigs
CC against parasitism from T. colubriformis.
CC See also P94260, and P94368-P94369.
SQ Sequence 320 AA;

Query Match 3.1%; Score 128; DB 1; Length 320;
Best Local Similarity 17.1%; Pred. No. 1.63e-01;
Matches 18; Conservative 35; Mismatches 49; Indels 3; Gaps 2;

Db 11 mkiekdnaldadaeekvrtgkierveeeldrtqkkmgtendlkagedlaatsq 70
QY 299 IEIROKEMLRNN--ESKIIHLQVIDSOAKELKELDKREIRPROMWEA-DSKMSVES 355
Db 71 leekexkvgaeaeavaalnrrmtlleeeleeraeerlkiateklee 115
QY 356 LQNRVTELESVDKSGAGVARNRTGLESLQSLRHDMLSVHDRLAD 400

RESULT 5
ID R66930 standard; Protein; 885 AA.
AC R66930;
DT 01-SEP-1995 (first entry)
DE AMML chromosome inv(16) product.
KW AMML; acute myelomonocytic leukemia; chromosome-16; inversion;
KW inv(16); CBF-beta; CBF gene; transcription factor; myosin; MYH11;
KW SMHC.
OS Homo sapiens.
FH Key
FH Peptide Location/Qualifiers
FT /label= CBF 1..164
FT Peptide 165..885
FT /label= MYH11
FN W09504067-A.
PD 09-FEB-1995.
PR 26-JUL-1994; U08530.
PR 29-JUL-1993; US-099869.
PA (UNMI) UNIV MICHIGAN
PA (TEXA) UNIV TEXAS SYSTEM.
PI Claxton D, Collins FS, Liu P, Siciliano MJ;
DR WPI; 95-082178/11.
DR N-PSDB; 084589.
PT Novel DNA spanning the pericentromeric inversion of chromosome 16 -
PT for the screening of acute myeloid leukaemia
PS Claim 4; Page 34-38; 78pp; English.
CC PCR was performed on total cellular RNA from 5 AMML patients having
CC a pericentromeric inversion of chromosome-16, M450 subtype. Sequencing
CC showed the inv(16) fusion to comprise a sequence from the CBF
CC gene, encoding a novel transcription factor, and the MYH11 gene,
CC encoding smooth muscle myosin heavy chain. In 1 patient, nt 1-492
CC of the CBF gene were fused to nt 994 of MYH11 (shown in

CC Q84589; predicted aa sequence in R66930). Probes based on inv(16)
CC can be used for diagnosis of AMML.
SQ Sequence 885 AA;

Query Match 3.1%; Score 130; DB 12; Length 885;
Best Local Similarity 22.2%; Pred. No. 1.17e-01;
Matches 28; Conservative 38; Mismatches 55; Indels 5; Gaps 5;

Db 204 leaqgqlskscdgeraraealndkvhlqnevsyrgmlneagkaiklakvaslssq 263
QY 269 LEKVSLQNSVEKKNKSISLHNQISFIEIERKEMLRNNESKIIHLQVIDSOAK 328
Db 264 lqdtqellqeetrklnvstklrg-leeernslqddqde-emeakqnlerhstlniqls 321
QY 329 LKELDKREIRP-FROMWEADSKMSVSLSQNRVTE-LESVDKSGQ-VARNRTGLESLQSL 385
Db 322 dekkkl 327
QY 366 RHDQML 391

RESULT 6
ID P81187 standard; protein; 1284 AA.
AC P81187;
DT 04-DEC-1990 (first entry)
DE Sequence encoded by a gene fragment coding for major protein of an A-type
DE inclusion body and a promoter region in poxvirus
KW Vaccinia virus; vaccine.
OS Cowpox virus.
FH Key
FH Region Location/Qualifiers
FT /note="Encoded by promoter region" 1..20
FN EP-261925-A.
PD 30-MAR-1988.
PF 22-SEP-1987; 308372.
PR 2-SEP-1986; JP-223972.
PR 09-SEP-1987; JP-223972.
PA (TOFU) Toa Nenryo Kogyo KK.
PI Shida H, Funahashi S;
DR WPI; 88-086185/13.
DR N-PSDB; n81538.
PT Gene fragment coding A-type inclusion body in pox virus -
PT used in the construction of recombinant vaccinia viruses for use
PT as vaccines
PS Disclosure; 3-10; 24pp; English.
CC The gene is nonessential for proliferation of poxvirus and is homologous
CC with a corresp. gene of a vaccinia virus and can therefore be used for
CC construction of a recombinant vaccinia virus for use as vaccines. The
CC promoter present upstream of this gene is very strong and is adequate as
CC promoter for expression of an exogenous antigen cell.
SQ Sequence 1284 AA;

Query Match 3.0%; Score 125; DB 1; Length 1284;
Best Local Similarity 22.4%; Pred. No. 2.66e-01;
Matches 22; Conservative 35; Mismatches 36; Indels 5; Gaps 5;

Db 673 lqsklsdlerlrcrnatelrqlg-yri-tdlerqlndcrnmenaa-dteremqrlr 729
QY 269 LEKVSLQNSVE-KKKSISLHNQISFIEIERKEMLRNNESKIIHLQVIDSOAK 326

Query Match 2.8%; Score 117; DB 11; Length 337;
Best Local Similarity 18.9%; Pred. No. 9.71e-01;
Matches 27; Conservative 47; Mismatches 62; Indels 7; Gaps 7;
Db 130 aeaillkvvklfgesrgenemekdrekadykknvdawgllreadkireanrlfav 189
QY 266 SNSLEKRVSLQNSVEKKNKSIQ-SLHNQCSFEIEIRKQEKMLHLOKVIDS 324
Db 190 nqkmtalekkkeaveegkrqientlkegnldideanrladeinslidyvediqtklppm 249
QY 325 QAEIKELDKKEIRF-RQNWELADSKMSVSESL-Q-NRVT-BLESDVSAGVANNITLL 380
Db 250 seelndkiddlsgelktrkjaek 272
QY 381 ESQLS-RHDDQ-LSVHDIRLADM 401

RESULT 10
ID P90994 standard; protein; 385 AA.
AC P90994;
DT 23-FEB-1990 (first entry)
DE B. subtilis sacuS1 polypeptide
KW Levan saccharase
OS Bacillus subtilis.
PN W08909264-A.
PD 05-OCT-1989.
PF 22-MAR-1989; F00134.
PR 22-MAR-1988; FR-003736.
PA (INSP) Inst Pasteur.
PI Kunst F, Debarbouille M, Msadek T, Rapoport G, Klier A, Decorder R, WPI; 89-309330/42.
DR N-PSDB; N91619
PT DNA contg. Bacillus subtilis sacu locus - for inducing overproduction in microorganisms
PS Claim 3; fig. 5; 87pp; French.
CC SacuS1 confers Degg phenotype, and is encoded by part of the 2.55 kb SalI-SphI Sacu locus. This can restore levan saccharase synthesis in CC B. subtilis sacu- mutants, and overproduces proteins in this or other microorganisms.
CC Sequence 385 AA;
SQ

Query Match 2.8%; Score 117; DB 1; Length 385;
Best Local Similarity 26.5%; Pred. No. 9.71e-01;
Matches 22; Conservative 30; Mismatches 27; Indels 4; Gaps 4;
Db 109 greqlreerddlerlllgig-elierseslvqitvlnylnqldrevglladagkq 167
QY 257 QHVLLEKWSLSLEKRVSLQNSVEKKNKSIQSLHNQCSFEIEIRKQEKML-RNNESKI 315
Db 168 dfgl-rileaegeerkrvsvreth 189
QY 316 -LHLQRIIDSOAEIKELDKKEIR 337

RESULT 11
ID R43563 standard; Protein; 476 AA.
AC R43563;
DT 05-APR-1994 (first entry)
DE Hyaluronan receptor.
KW Hyaluronan binding protein; HA; RHAMM; mediated motility; wound;

KW healing; diagnosis; treatment; cell locomotion; tumour invasion;
KW birth defects; inflammatory disorder; Alzheimer's disease; dementia;
KW Parkinson's disease; Huntington's disease; AIDS; diabetes; auto;
KW immune diseases; corneal dysplasia; hypertrophy; surgery; burns;
KW stroke; multiple sclerosis; depression; schizophrenia; CNJ;
KW contraception; in vitro fertilisation; embryo development.
FN W09321312-A.
PD 28-OCT-1993.
PF 13-APR-1993; CA0158.
PR 09-APR-1992; GB-007949.
PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
PA (DTMA-) UNIV.
PA MANITOBA.
PI Turley EA;
DR WPI; 93-351722/44.
DR N-PSDB; 051212.
PT DNA encoding hyaluronan receptor - used to produce proteins and antibodies for alteration of cell locomotion
PS Claim 7; Fig 23; 88pp; English.
CC The sequence is that encoded by a cDNA clone encoding the hyaluronan receptor (HARC). The sequence was obt. by screening a 373 library in lambda gtl with antibodies to HARC. A clone of 1.9 kb was obtained and used to rescreen the library to obtain the full length, 2.9 kb clone. HA is down regulated in stationary normal cells and is only expressed in situations where cell motility is desired, e.g. in wound healing, in response to growth factors and in chemotaxis by white blood cells. HA may be used for diagnosis and treatment of diseases involving cell locomotion, e.g. tumour invasion, birth defects, acute and chronic inflammatory disorders, Alzheimer's and other forms of dementia, AIDS, diabetes, autoimmune diseases, corneal dysplasias and hypertrophies, burns, surgical incisions and adhesions, CC strokes, multiple sclerosis, depression/schizophrenia related to CC neuronal growth and pain states involving nerve sprouting; also in CNJ CC and spinal cord regeneration, contraception, in vitro fertilisation and CC embryo development.
CC See also R46548-51.
CC Sequence 476 AA;
SQ

Query Match 2.8%; Score 118; DB 8; Length 476;
Best Local Similarity 20.9%; Pred. No. 8.27e-01;
Matches 31; Conservative 41; Mismatches 71; Indels 5; Gaps 5;
Db 101 nlirekevelekhiaargaqailagkyidtaga-lrvvtaqlesvqekyndtaglvdv 159
QY 260 NLKEMWSLSLEKRVSLQNSVEKKNKSIQSLHNQCSFEIEIRKQEKMLRNNESKIHL 318
Db 160 taqlesvqekyndtaglrvvtaqlesvqekyndtaglrvvtaqlesvqekyndtagl 219
QY 319 QRVIDSQAEIKELDKKEIRFQNWELADSKMSVSESLQNRVTELESDVSAGVANNIT 377
Db 220 rdvasqlesyksstlkeidklentl 247
QY 378 GLIESQLSRH-DQMS-VHDIRLADM 403

RESULT 12
ID R05766 standard; protein; 462 AA.
AC R05766;
DT 05-NOV-1990 (first entry)
DE Portion of peptide antigen to malarial sporozoite.

Query Match	2.7%;	Score 114;	DB 1;	Length 462;
Best Local Similarity	21.9%;	Pred. No. 1.57e+00;		
Matches	32;	Conservative	46;	Mismatches 60; Indels 8; Gaps 8;

```

Dh 204 lakek1qegqsd1egger1kek1qegq-s-d1egdr1akek1qegqsd1egdr1akek1q 261
    | | | | | : : : : : | : | | | | : : : : : | : | | | | : : : : :
Qy 269 LER-KO-SLLONEVEKKNKSIOSLNHOJCSFETELER-OKMLNNEKSLIHLIÖR1IDSQA 326
    | | | | | : : : : : | : | | | | : : : : : | : | | | | : : : : :
Dh 262 egqsd1egdr1akek1qvgqsd1ertkastet1qegrsd1egger1kek1qegqsdl-e-q 320
    | : | : : : : : : : : : | : | | | | : | : | : : : : | : |
Qy 327 EKLKEKLDKE-I-RP-FRÖMWEZADSMKSSVESLÖNRVTELESVDKSAGOVANRTGLESO 383
    | : | : : : : : : : : : | : | | | | : | : | : : : : | : |
Dh 321 errakek1qegqr1qeger1akek1q 346
    | : | : : : : : : : : : | : | | | | : | : | : : : : | : |
Qy 384 LSRFDQMLSVHDIRLADMDLGRFYLE 409

```

ID	RESULT	13
AC	P90954	standard; protein; 769 AA.
DT	27-Feb-1990	(first entry)
DE	Yeast topoisomerase I cDNA	
KW	Scleroderma.	
OS	Saccharomyces cerevisiae	
PN	MO8909222-A.	
PD	05-OCT-1989.	
PF	22-MAR-1989;	001116.
PR	23-MAR-1988;	US-172159.
PA	(BBIG) Brigham and Women's Hospital;	(UYJO) John's Hopkins Univ.
PI	Earnshaw WC, D'Arpa P;	
DR	WPI; 89-309500/42.	
PT	Cloned cDNA encoding eukaryotic topoisomerase I -	useful for large scale
PT	prodn. by recombinant methods	
PS	Disclosure; fig. 6; 28pp; English.	
CC	The cDNA of this can be spliced into DNA vectors and used to transform	
CC	hosts for high yield. This has high homology with human topoisomerase I	
Seq	sequence	769 AA;
Query Match	2.7%;	Score 111; DB 1; Length 769;

Best Local Similarity 24.7%; Pred. No. 2.52e+00;
Matches 23; Conservative 26; Mismatches 41; Indels 3; Gaps 3;

RESULT	14	Standard; Protein; 870 AA.
ID	R30729	
AC	R30729;	
DT	20-MAY-1993	(first entry)
DE	p100 protein from human herpes virus type 6.	
KW	antibody's; monoclonal antibody; ELISA assays; CMV; cytomegalovirus.	
OS	Human herpes virus type 6.	
PN	EP-524421-A.	
PD	27-JAN-1993.	
PF	15-JUN-1992; 110047.	
PR	08-JUL-1991; EP-111338.	
PA	(BEHW) BEHRINGER AG.	
PI	Fleckenstein B, Neipel F;	
DR	WPI; 93-028531/04.	
PT	P-PDSB; R30729.	
PT	Human herpes virus type 6 protein p100 DNA sequence - useful in	
PT	prophylaxis, treatment and differential diagnosis of human herpes	
PT	virus-6 infections	
FS	Claim 1; Page 12; 25pp; English.	
CC	This sequence is the p100 protein from human herpes virus type6.	
CC	The protein and antibodies to it can be used for treatment or	
CC	prevention of HHV-6 infections. The DNA , protein and Ab are also	
CC	useful in eg. ELISA assays esp. for differentiating between HHV-6	
CC	and cytomegalovirus infections. These assays are more sensitive	
CC	and specific than immunofluorescence methods currently used.	
SO	Sequence 870 AA;	

Query Match	2.7%	Score 111; DB 6; length 870;
Best Local Similarity	26.4%	Prod. No. 2.52e+00;
Matches	32; Conservative	34; Mismatches 47; Indels 8; Gaps 7
D6	678	gryvmlisqfnnlslsgekgiedlqn-qrgtelklatenksgreseseanvekil-eysn 735
Oy	319	QRRVD--SQAEIKEL-DKEIIPFQRMEEFDSKRSVESLQNRVTELSVSKAGVARR 375
D6	736	pqdmfkn-fllqncldwsgsflrpdadls-reldsaafcdaldiklpgngereidale 793
Oy	376	NTGLIESQLSRHDOMLSVHD-IRLDMDLGRQVLETRSYNGVLTKIKIRYKRRKQEAVMG 434
D6	794	K 794
Oy	435	K 435

RESULT 15

ID R15148 standard; Protein; 475 AA.

AC R15148;

DT 14-FEB-1992 (first entry)

DE Ro/SSA autoantigen.
 KM Autoantibody; autoantigen; SLE; systemic lupus erythematosus.
 OS Homo sapiens.
 PN M09117171-A.
 PD 14-NOV-1991.
 PF 07-MAY-1991; U03139.
 PR 07-MAY-1990; US-520270.
 PA (OKLA-) OKLAHOMA MED RES FO.
 PI Frank MB, Itoh K.
 DR WPI; 91-355732/48.
 DR N-PDB; Q14798.
 FT DNA encoding an Ro-SSA autoantigen - useful for diagnosing
 PT auto-immune disorders or presence of auto-antibodies
 PT Disclosure; Fig 2; 41pp; English.
 CC A cDNA library (from human thymus mRNA) in lambda gII was screened
 CC with serum from a patient having systemic lupus erythematosus) Two
 CC clones were reactive with sera (from a panel of lupus patients)
 CC which contd. autoantibodies against 52 KD protein.
 CC Both the cDNA and the protein expressed from it, or portions of it,
 CC are useful as diagnostic agents in the identification of patients
 CC having autoantibodies and in the identification and analysis of
 CC the structural and functional properties of the autoantigen and for
 CC application in immunotherapeutic regimens.
 QJ Sequence 475 Aa;

Query Match	2.6%;	Score 107;	DB 3;	Length 475;
Best Local Similarity	27.3%;	Pred. No. 4.70e+00;		
Matches	12;	Conservative	13;	Mismatches 19;
				Indels 0;
				Gaps 0;

```

Db      16 cpicldfrfvepvsiecghsfcgacisgvkrkggsvcpvcrqrfi 59
      | | : | ||| || : : : : | | : :
Gy      53 CEKHLVLCSPQTECGHRCESMAALLSSSPKCTACQESIV 96

```

Search completed: Tue Mar 19 09:53:45 1996
Job time : 28 secs.

Release 2.1.D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

```

MPsrch_gp    protein - protein database search, using Smith-Waterman algorithm
Run on:      Tue Mar 19 09:52:04 1996;      MacPar time 18.51 Seconds
Tabular output not generated.              775.963 Million cell updates/sec
Title:       >US-08-404-832-2

```

Description:	(1-567) from US08404832.pep
Perfect Score:	4162
Sequence:	1 MESSKMDSPGALQINPPLK.....IKDDTIFIKVIYDTSLEDP 567

Scoring table: PAM 150
Gap 11

Searched: 82306 seqs, 25270970 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

```
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4
8:unann5 9:unann6 10:unann7 11:unann8 12:unrev1
13:unrev2
```

Statistics: Mean 49.231; Variance 140.106; scale 0.351

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No
1	4139	99.4	368	11	A53649	TNR-associated prot	0.00e+00
2	812	19.5	416	11	B55649	TNR-associated prot	2.29e-93
3	191	4.6	458	11	A23361	DG17 protein - alime	9.46e-09
4	164	3.9	760	11	A04195	meprin A (EC 3.4.24.	1.10e-05
5	149	3.6	283	5	A60364	tropomyosin - mbgrt	4.63e-04
6	151	3.6	377	11	A47380	RING finger-containi	2.84e-04
7	151	3.6	748	11	S24134	endopeptidase 2 (EC	2.84e-04
8	151	3.6	1969	2	S02771	myosin heavy chain A	2.84e-04
9	145	3.5	284	5	A44980	tropomyosin, oblique	1.23e-03
10	145	3.5	1959	2	A33977	myosin heavy chain.	1.23e-03
11	141	3.4	1961	2	A61231	myosin heavy chain N	3.21e-03
12	142	3.4	1966	2	MGKW	myosin heavy chain B	2.53e-03
13	132	3.2	284	5	S24972	tropomyosin alpha, c	2.67e-02
14	132	3.2	527	10	S33068	myosin II heavy chai	2.67e-02
15	132	3.2	700	10	S49383	meprin A (EC 3.4.24.	2.67e-02
16	135	3.2	714	13	S34644	PABA peptide hydrola	1.33e-02
17	135	3.2	1979	2	S03166	myosin heavy chain.	1.33e-02
18	135	3.2	3259	10	S37536	macrogolgin - human	1.33e-02
19	130	3.1	221	5	S03446	tropomyosin, nonmusc	4.24e-02
20	130	3.1	245	13	S24043	tropomyosin alpha -	4.24e-02
21	128	3.1	248	5	D34787	tropomyosin 3 alpha,	6.70e-02
22	128	3.1	245	5	D39816	tropomyosin 5b, fibr	6.70e-02
23	128	3.1	248	5	S39816	tropomyosin 5a, fibr	6.70e-02
24	128	3.1	251	5	B34787	tropomyosin 2 alpha,	6.70e-02
25	131	3.1	227	5	A02991	myosin heavy chain,	3.37e-02
26	128	3.1	284	13	S23470	tropomyosin - Atrica	6.70e-02
27	131	3.1	465	5	A02986	myosin alpha heavy c	3.37e-02
28	129	3.1	944	9	S26710	spindlin pole body pr	5.33e-02
29	127	3.1	961	5	A33626	fibrinogen alpha cha	8.41e-02
30	127	3.1	1053	10	A41642	dynactin - chicken	8.41e-02
31	127	3.1	1078	5	A30220	myosin heavy chain,	5.33e-02
32	128	3.1	1132	9	S37932	hypothetical protein	6.70e-02

33	131	3.1	1509	2	A27224	myosin heavy chain I	3.37e-02
34	129	3.1	1938	2	MMKWL	myosin heavy chain D	5.33e-02
35	130	3.1	1998	2	S21801	myosin heavy chain, 4.24e-02	
36	127	3.1	2007	2	B43402	myosin heavy chain-B	8.41e-02
37	130	3.1	2094	11	S33124	tpi protein - human	4.24e-02
38	129	3.1	2116	5	A26655	myosin heavy chain -	5.33e-02
39	126	3.0	248	5	B30125	tropomyosin beta-3,	1.06e-01
40	126	3.0	621	10	A29965	lamin Dm-0 precursor	1.06e-01
41	126	3.0	676	5	S00084	myosin heavy chain,	1.06e-01
42	125	3.0	1325	13	S16129	dysen-associated pr	1.32e-01
43	126	3.0	1437	9	S12459	myosin beta heavy ch	1.06e-01
44	126	3.0	1679	9	S48385	hypothetical protein	1.06e-01
45	126	3.0	1937	10	S49478	myosin - human	1.06e-01

ALIGNMENTS

RESULT 1
ENTRY A55649 #type complete
TITLE TNFR-associated protein LMP1 - human
ALTERNATE_NAMES CD40-binding protein
ORGANISM #formal name Homo sapiens #common name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 03-Apr-1995

ACCESSIONS
REFERENCE A55649
#authors Mosialos, G.; Birkenbach, M.; Valamanchili, R.; VandeRade, T.; Mare, C.; Kieff, E.

#journal Cell (1995) 80:389-399
#title The Epstein-Barr virus transforming protein LMP1 engages signaling proteins for the tumor necrosis factor receptor family.

#accession A55649
#molecule_type mRNA
#residues 1-568 #label MOS
#cross-references GB:019260

REFERENCE A55135
#authors Hu, H.M.; O'Rourke, K.; Boguski, M.S.; Dixit, V.M.
#journal J. Biol. Chem. (1994) 269:30069-30072
#title A novel RING finger protein interacts with the cytoplasmic domain of CD40.

#accession A55135

#molecule_type mRNA
#residues 1-133,135-404,'G',406-568 #label HDA
#cross-references GB:015637

KEYWORDS
FEATURE #note coiled coil
#note nucleotide sequence not given

SUMMARY
#region RING-finger motif
#length 568 #molecular-weight 64490 #checksum 8660

Query Match 99.4%; Score 4139; DB 11; Length 568;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 566; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 1 mesekndspalgtluplkthdrasgtvfvpeggykktvktvckvckchyl 60
1 MESKKKDDSPALGTLPKLTHTDRSGTFFVPEGGYKKTVKTVCKCKCHYL 60

Db	61	csptqteqhnfrceesmaallssespkctacgesiykdvfkdnockreialdiqyne	120
Qy	61	CSPTQTEQHFRCEESMAALLSSSPKCTACGESIYKDVFKDNCKREIALDIQYCN	120
Db	121	argaeqqlmghllyhlnkchfeelpcyvppdkvllkldrhvckxyreacahc	180
Qy	121	ARGAEQQLMGLLHLYHKNCHFEELPCYVPDPKVLKLDLHVEKAKYREATCSHC	179
Db	181	ksqymalqkhdtdpcvvascpnkcsvqtlireelsahlsecvnapstcsfrkycv	240
Qy	180	KSQYPMALQKHEDTDCPCVVASCPHKCSVQTLIRELSAHLSECVNAPSTCSFRKCV	239
Db	241	fgtngqikahaeasavqhvnlkewanslekksvllqnesveknksiqslmqicse	300
Qy	240	FGTNGQIKAHAEASAVQHVNLKEWNSLEKKSILLQNESEKKSIIQSLNQICSEI	299
Db	301	eierqemlrneekihlgrrvidsqeekikeidkeirfrqmwesdamkssveslqr	360
Qy	300	EIERQEMLRNESKILHLGRRVIDSQEKIKELDKELRFRQWESADSKSSVESLQNR	359
Db	361	vtelssvdksagvvarntglleeqlerhdqmlsvhdirladnrlrfvletasyngvliw	420
Qy	360	VTELESVDKSGAVARNTGLLEEQLSRHDQMLSVHDIRLADNRLRFVLETASVNGVLIW	419
Db	421	kirdykrkqgaumgtclsiysqpfyfygkmcarylmgdmqgkgtahlsffvrmg	480
Qy	420	KIRDYKRKQGAUMGKTLISYSPFYGYGKMCARYLMDGDMGKGTHTLSFFVIRMG	479
Db	481	eydallpwpfkykvrlmldmggsrhlhgafkfpdpnasfkkprgmmlaagcvfvag	540
Qy	480	EYDALLPWPFKYKVRLLMDGSSRHLHGAFKFPDPNSSFFKPRTEAMIASGCVFVAQ	539
Db	541	tvlenqykddtffikvltedslpdp	568
Qy	540	TVLENGYIKDDTFFIKVIVDTSLDPDP	567

RESULT 2
ENTRY B55649 #type complete
TITLE TNFR-associated protein EB16 - human
ORGANISM #formal name Homo sapiens #common name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 23-Mar-1995

ACCESSIONS
REFERENCE B55649
#authors Mosialos, G.; Birkenbach, M.; Valamanchili, R.; VandeRade, T.; Mare, C.; Kieff, E.

#journal Cell (1995) 80:389-399
#title The Epstein-Barr virus transforming protein LMP1 engages signaling proteins for the tumor necrosis factor receptor family.

#accession B55649
#molecule_type mRNA
#residues 1-416 #label MOS
#cross-references GB:019261

SUMMARY
#length 416 #molecular-weight 46163 #checksum 6815
Query Match 19.5%; Score 812; DB 11; Length 416;
Best Local Similarity 38.9%; Pred. No. 2.29e-93;

ENTRY	5	#type complete
TITLE	A60364	tropomyosin - migratory locust
ORGANISM	#formal_name Locusta migratoria	#common_name migratory locust
DATE	03-Feb-1993	#sequence_revision 03-Feb-1993
ACCESSIONS	A60364	18-Jun-1993
REFERENCE	A60364	Krieger, J.; Ramling, K.; Knipper, M.; Grau, M.; Mertens, S.; Beer, H.
#journal	Insect Biochem. (1990) 20:1173-184	
#title	Cloning, sequencing and expression of locust tropomyosin.	
#accession	A60364	
#status	not compared with conceptual translation	
##molecule_type	mRNA	
##residues	1-283	##label KRI
CLASSIFICATION	#superfamily tropomyosin	
KEYWORDS	coiled coil; hepled repeat	
SUMMARY	#length 283; #molecular-weight 32439	#checksum 4917
Query Match	3.6%; Score 149; DB 5; Length 283;	
Best Local Similarity	19.0%; Pred. No. 4,63e-04;	
Matches	20; Conservative	36; Mismatches 46; Indels 3; Gaps 3;
Db	13 lekdnaldrallceqgdanl-ra-ekaeearalqkktigtendldqtgeslqymak 70	
Oy	297 FEIEIRKEMLRNNESEKILRLORVIDISOAEKLEKDEKIRPFROMWEA-DSMKSYES 355	
Db	71 leekelagaaeseaalnrridgleedlargeelataataakae 115	
Oy	356 LQNRVTELESYDKSAGVARNRTGILLESQLSRHDQMLSVHDRIAD 400	
RESULT	6	
ENTRY	A47380	#type complete
TITLE	RING finger-containing DNA binding protein RING1 - human	
ORGANISM	#formal_name Homo sapiens	#common_name man
DATE	21-Jan-1994	#sequence_revision 18-Nov-1994
ACCESSIONS	A47380	18-Nov-1994
REFERENCE	A47380	Lowering, R.; Hanson, I.M.; Borden, K.L.; Martin, S.; O'Reilly, N.J.; Evan, G.I.; Rahman, D.; Pappin, D.J.; Trowadeale, J.; Freemont, P.S.
#author	Proc. Natl. Acad. Sci. U.S.A. (1993) 90:2112-2116	
#journal	Identification and preliminary characterization of a protein	
#title	motif related to the zinc finger.	
#cross-references	MUID:93211912	
#accession	A47380	
##status	preliminary	
##molecule_type	DNA; protein	
##residues	1-377	##label LOV
##cross-references	NCBIN:128010; NCBIPI:128011	
#note	sequence extracted from NCBI backbone	
SUMMARY	#length 377	#molecular-weight 39145
		#checksum 7895
Query Match	3.6%; Score 151; DB 11; Length 377;	
Best Local Similarity	24.7%; Pred. No. 2.84e-04;	
Matches	24; Conservative	30; Mismatches 39; Indels 4; Gaps 4;

Db	11	relhnelmelpldmlkntmtcklrfedscivnlalrsgmkpeptckklivkrsrlrp	70
Oy	45	KIVEDKTKKCKEHWLVCSPKOT-ECGHRFCESCAALISSSPKCTACQESIVKDKVKRD	103
Db	71	dp-nfdalisklypsreeyeahqdrvl-i1srlhng	105
Oy	104	NCKRETLALQIY-CRNESRCGACQIMLGHVHKND	139
RESULT	7		
ENTRY		S24134	#type complete
TITLE		endopeptidase 2 (EC 3.4.24.-) - rat	
ALTERNATE_NAMES		endopeptidase 24.18	
ORGANISM		#formal name Rattus norvegicus	#common name Norway rat
DATE		02-Dec-1993	#sequence_revision 01-Sep-1995
ACCESSIONS		01-Sep-1995	#text_change
REFERENCE		S24134	
#authors		S24134	
#journal		Corbell, D.; Gaudoux, F.; Wainwright, S.; Ingram, J.; Kenny	
#title		A.J.; Boileau, G.; Crine, P.	
		FEBS Lett. (1992) 309:203-208	
		Molecular cloning of the alpha-subunit of rat	
		endopeptidase-24.18 (endopeptidase-2) and co-localization	
		with endopeptidase-24.11 in rat kidney by in situ	
		hybridization.	
		hybridization.	
#cross-references		MFID:92371675	
#accession		S24134	
#status		preliminary	
#molecule		type mRNA	
#residues		1-748	#label COR
KEYWORDS		hydroxylase; metalloprotein; proteinase; zinc	
FEATURE			
156,160,166			
SUMMARY		#binding site zinc (His) #status predicted	
		#length 748 #molecular-weight 85138	#checksum 2333
Query Match		3.6%; Score 151; DB 11; Length 748;	
Best local similarity		28.9%; Pred. No. 2,84e-04;	
Matches		26; Conservative 23; Mismatches 37; Indels 4; Gaps 4;	
Db	436	wctirnisqvlentvktgdl-v-sprfyne-ygfgfvtllypngritsnsgylglaflny	492
Oy	418	IKRIIRYKRRKKEAWGKTLISYSPFTYTGFGKMCARVYINGDMGKTHLSLFVIM	477
Db	493	sgndrvllwpyvenegaimtlldgdpdarn	522
Oy	478	RGEYDALILPFRKQKVITIM-LMDGSSSRH	506
RESULT	8		
ENTRY		S02771	#type complete
TITLE		myosin heavy chain A - Caenorhabditis elegans	
CONTAINS		myosin ATPase (EC 3.6.1.32)	
ORGANISM		#formal name Caenorhabditis elegans	
DATE		31-Dec-1993	#sequence_revision 31-Dec-1993
ACCESSIONS		27-Jan-1995	#text_change
REFERENCE		S02771	
#authors		Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.	
#journal		J. Mol. Biol. (1989) 205:603-613	
#title		Sequence analysis of the complete Caenorhabditis elegans	

Db	85	vaalnrrmtlleeleraeerikatekleeatnhd-eeseryrk-menisfgde-era	141
Qy	287	IOSHHQHSFTEIEEROKEMRNNSKILHLOVIDSQAELKELEDRPFRRNMEFA	346
Db	142	ntleaglkqad-mlae-ea-dtkydevarklamreadleraaeraaagenkiveleelr	198
Qy	347	DSKSSVESLQNRVTELESVDKSAGQVARNRTGLIESQLSRHDMISVHDRIADHDLFQ	406
Db	199	vv	200
Qy	407	VL	408
RESULT	10		
ENTRY	A33977	#type complete	
TITLE		myosin heavy chain, cellular - chicken	
CONTRAINS		myosin ATPase (EC 3.6.1.32)	
ORGANISM		#formal name Gallus gallus #common name chicken	
DATE		31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 27-Jan-1993	
ACCESSIONS	A33977		
REFERENCE		Shohei, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S. Proc. Natl. Acad. Sci. U.S.A. (1989) 86:7726-7730	
#journal		Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular myosin.	
#title		vertebrate cellular myosin.	
#cross-references	M01D:9004668		
#accession	A33977		
##residues	1-1959	##label SHO	
##molecule_type	mRNA		
##cross-references	GB:W26510		
CLASSIFICATION		#superfamily myosin heavy chain; myosin head homology actin binding; ATP; coiled coil; hydrolase; methylation; tandem repeat	
KEYWORDS			
FEATURE			
84-1764		#domain myosin head homology #label HEA\	
174-181		#region nucleotide-binding motif A (P-loop)\	
552-563		#region actin-binding #status predicted\	
626-640		#region actin-binding #status predicted\	
837-1936		#domain coiled coil #status predicted #label COI\	
837-1277		#region S2\	
1278-1959		#region light meromyosin\	
1937-1959		#domain carboxyl-terminal #label CBT\	
125		#modified site N6,N6,N6-trimethyllysine (Lys) #status predicted\	
180		#binding site ATP (Lys) #status predicted\	
694, 704		#active site Cys #status predicted	
SUMMARY		#length 1959 #molecular-weight 226502 #checksum 3641	
Query Match	3.5%	Score 145; DB 5; Length 284; Mismatches 41; Indels 6; Gaps 6;	

RESULT	ENTRY	TITLE	CONTRINS	ORGANISM	DATE	ACCESSIONS	REFERENCE
11	A61231	#type complete myosin heavy chain NMHC-A	nonmuscle - human	myosin ATPase (EC 3.6.1.32)	#normal_name Homo sapiens #common_name man 12-May-1994 #sequence_revision 14-JUL-1994 #text_change 27-Jan-1995	A61231; A34876	A61231

#authors	#journal	#title
Simons, M.; Wang, J.; McBride, O.W.; Kawanoto, S.; Yamakawa, K.; Gadula, D.; Adelstein, R.S.; Weir, L.	Circ. Res. (1991)	69:530-539

Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes.

```
##accession      Ab1231
##molecule_type mRNA
##residues      1-715 ##label SIM
##cross-references GB:M69180
REFERENCE
A34876
```

#authors Saez, C.G.; Myers, J.C.; Snows, T.B.; Lelmand, L.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:1164-1168
#title Human nonmuscle myosin heavy chain mRNA: generation of diversity through alternative polyadenylation.
#cross-references PMID:90138958
#accession A34876

```
##molecule_type mRNA
##reichers 715-1961 ##label SAE
##cross-references GB:M31013
GENETICS
```

#gene CDB:MYH9
#map position 22q12.3-q13.1
CLASSIFICATION #superfamily myosin heavy chain; myosin head homology
KEYWORDS actin binding; ATP; coiled coil; hydrolase; methylation;
tandem repeat

FEATURE	
84-164	#domain myosin head homology #label HEA\
152-161	#region nucleotide-binding motif A (P-loop)\
574-585	#region actin-binding #status predicted\
626-640	#region actin-binding #status predicted\
837-1938	#domain coiled coil #status predicted #label COI\
837-1277	#region S2\
1278-1961	#region light meromyosin\
1939-1961	#domain carboxyl-terminal #label CBT\
125	#modified_site N6,N6,N6-trimethyllysine (Lys) #status

RESULT	12
ENTRY	MMKM
TITLE	#type complete
CONTRIS	myosin heavy chain B - Caenorhabditis elegans
ORGANISM	myosin ATPase (EC 3.6.1.32)
DATE	#formal_name Caenorhabditis elegans
	13-Jun-1983 #sequence_revision 19-Feb-1984 #text_change
	27-Jan-1995

ACCESSIONS	REFERENCE
A93956; A93287; A02992	Karn, J.; Brenner, S.; Barnett, L. Proc. Natl. Acad. Sci. U.S.A. (1983) 80:4253-4257
A93956	Protein structural domains in the <i>Caenorhabditis elegans</i>

```

unc-34 myosin heavy chain gene are not separated by
introns.
#cross-references MWID:83673600
#accession A93958
##molecule type DNA

```

##residues	1-1966	##label	KAR
##cross-references	GB:J01050		
REFERENCE			
##authors	A93287		
##journal	McLachlan, A.D.; Karn, J.		
##title	Nature (1982) 299:226-231		
	Periodic charge distributions in the myosin rod amino acid		

```
sequence match cross-bridge spacings in muscle.
#cross-references MUID:82272395
#accession A93287
##molecule type DNA
```

GENETICS	##residues	850-1336,'R',1338-1873,'L',1881-1966	##label	MCL
##gene				
##introns				
CLASSIFICATION				
	unc-54			
	21/3;	64/2;	114/3;	267/1;
				528/3;
				1750/3;
				1822/3;
				1897/3
	##superfamily myosin heavy chain; myosin head homology			

KEYWORDS	FEATURE
actin binding; ATP; coiled coil; hydrolase; methylation; muscle contraction; tandem repeat	#domain myosin head homology #label HEA\ #region nucleotide-binding motif A (p-loop)\ #region actin-binding #status predicted\ #region actin-binding #status predicted\ #domain coiled coil #status predicted #label COI\ #region S2\ #region light meromyosin\ 87-778 177-184 665-687 769-783 851-1966 851-1165 1166-1966

	128	#modified site N6,N6,N6-trimethyllysine (Lys) #status predicted\
	183	#binding site ATP (Lys) #status predicted\
SUMMARY	705,715	#active site Cys #status predicted #length 1966 #molecular-weight 225125 #checksum 6382
Query Match		3.4% Score 142; DB 2; Length 1966;
Best Local Similarity	24.2%;	Pred. No. 2.53e-03;
Matches	32; Conservative	37; Mismatches 53; Indels 8; Gaps 8;
Db	984 dhnqirsldemgqqdaiaiklnk-e-kheeinklmedjqseedkgbnqnkvakleqt	1042
Oy	270 EKKVSILOÑSVEKNKSIOISLNQICSPFEIERQ-KEMIRNNESKILHLGVDSQAELK	328
Db	1043 lddledslerekarardldkqrkveg-elkiaq-enides-gr-qrh-dl-emllkke	1036
Oy	329 ILEKDKEIRPFRRONWEADSMKSVSESLONNVTELESVDKSGAGAVANTGLESLSRD	388
Db	1097 selhvassrlrd	1108
Oy	389 QMLSVDIRLAD	400
RESULT	13	
ENTRY	S24972	#type complete
TITLE	tropomyosin alpha, cardiac - pig	
ORGANISM	#formal_name Sus scrofa domestica #common_name domestic pig	
DATE	20-Feb-1995	#sequence_revision 20-Feb-1995 #text_change 12-Apr-1995
ACCESSIONS	S24972	
REFERENCE	S24972	
#author	Whitby, F.G.; Kent, H.M.; Stewart, F.; Stewart, M.; Xie, X., Hatch, V.; Cohen, C.; Phillips Jr., G.	
#submission	submitted to the EMBL Data Library, April 1992	
#description	Structure of tropomyosin at 9 Angstroms resolution.	
#accession	S24972	
#status	Preliminary	
#molecule_type	mRNA	
#residues	1-284	#label WHI
CLASSIFICATION	#cross-references EMBL:X66274	
SUMMARY	#superfamily tropomyosin	
Query Match		3.2% Score 132; DB 5; Length 284;
Best Local Similarity	23.9%;	Pred. No. 2.67e-02;
Matches	27; Conservative	39; Mismatches 39; Indels 8; Gaps 6;
Db	45 alqkkikatedelkysealkdageklelaekkatdaead-vaslrrriqlf-e-weeld	100
Oy	289 SLHNQICSPFEIERQKEMIRNNESKI-LHLPVIDSOAKELKELDKERIPFRONWEAD	347
Db	101 --taq-erlatatqlkeeeakaadesergmkvvisrzdqkdeekmeiqeiqlke	150
Oy	348 SMKSVSESLONNVTELESVDKSGAGAVANTGLESLSRDOMLSVDIRLAD	400
RESULT	14	
ENTRY	S33068	#type complete
TITLE	myosin II heavy chain - fluke (Schistosoma mansoni)	
ORGANISM	#formal_name Schistosoma mansoni	

DATE	22-Nov-1993 #sequence_revision 26-May-1995 #text_change 26-May-1995
ACCESSIONS	S33068
REFERENCE	A46514
#authors	Solisson, L.M.A.; Masteron, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.
#journal	J. Immunol. (1992) 149:3612-3620
#title	Induction of protective immunity in mice using a 62-kDa recombinant fragment of a Schistosoma mansoni surface antigen.
#cross-references	MOTID:J93056536
#accession	S33068
##status	preliminary
##molecule_type	mRNA
##residues	1-527 ##label SOI
##cross-references	EMBL:X65351
##note	the authors translated the codon CAA for residue 346 as lys
SUMMARY	#length 527 #molecular-weight 61535 #checksum 5188
Query Match	3.2%; Score 132; DB 10; Length 527;
Best Local Similarity	24.4%; Pred.No. 2.67e-02;
Matches	29; Conservative 33; Mismatches 51; Indels 6; Gaps 4,
D6	255 dneurtlgseamagdegmglkdkdklneenkrtglaqaeedkvnhlkakleest 314 ::: :
Oy	270 EKVSLLQNSVEKKNKSLOSLSHQCSEFIEIKQKEMLRNNSSKLILHQRTDSQA 329 ::: :
D6	315 demeenlareqtkirgvekskrleg-dlkatg-etvd-dlervkrd--leeqlrke 367 : :: :
Oy	330 KEIDKEIRPROMWEADSKMSVSESLONRVTELESVDKSGCVAHNTLLESQLSRD 388 : :: :
RESULT	15
ENTRY	S49383 #type complete
TITLE	meprin A (EC 3.4.24.18) - human
ORGANISM	16-Feb-1995 Homo sapiens #common name man
DATE	26-May-1995 #sequence_reviseion 12-May-1995 #text_change
ACCESSIONS	S49383
REFERENCE	S49383
#authors	Eldering, J.A.; Groenberg, J.; Sterchi, E.E.
#submission	submitted to the EMBL Data Library, September 1994
#description	Cloning and the PDBA-peptide hydrolase beta subunit: coexpression is required for plasma membrane localization of the alpha subunit in COS-1 cells.
#accession	S49383
##status	preliminary
##molecule_type	mRNA
##residues	1-700 ##label EMD
##cross-references	EMBL:X61333
CLASSIFICATION	#superfamily MAM homology; astacin homology
FEATURE	
11-257	#domain astacin homology #label AST\
260-429	#domain MAM homology #label MAM
SUMMARY	#length 700 #molecular-weight 79458 #checksum 916
Query Match	3.2%; Score 132; DB 10; Length 700;
Best Local Similarity	33.3%; Pred. No. 2.67e-02;
Matches	30; Conservative 20; Mismatches 29; Indels 11; Gaps 7;

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 [1]
 RP SEQUENCE FROM N.A.
 RM 94349371
 RA ROTHE M., WONG S.C., HENZEL W.J., GOEDEL D.V.;
 RL CELL 78:681-692(1994).
 CC -1- FUNCTION: SIGNAL TRANSDUCER ASSOCIATED WITH THE CYTOPLASMIC DOMAIN
 OF THE 75 KD TUMOR NECROSIS FACTOR RECEPTOR (TNF-R2).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SUBUNIT: HETERODIMER OF TRAF1 AND TRAF2.
 CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
 DR EMBL; L33303; MTRAF2A.
 KW ZINC-FINGER; COILED COIL.
 FT ZN_FING 34 72 C3HC4-TYPE.
 SQ SEQUENCE 501 AA; 56026 MM; 1264825 CN;
 Query Match 19.2%; Score 798; DB 7; Length 501;
 Best Local Similarity 52.7%; Pred. No. 5.03e-123;
 Matches 98; Conservative 49; Mismatches 37; Indels 2; Gaps 2;
 Db 317 lsntvqlersjgkldiamdeqkvelevstvgvfiwkadfrkxgeavgrtai 376
 QY 380 LESQLSRHDQMLSVHDIRLADMDIGFOVLETAISINGVLWKIDYRRKQEVNMGKTI 439
 Db 377 fspafyerygkncilvylngdgtgrghlsifvnmkypndallqwpfnqkvlmld 436
 QY 440 YSOFYTGFGYKRCARVYLNQDGMKGTHLSLFVIMRGEVDALLPWFKQVLTMLMD 499
 Db 437 h-nurehaidafpdrsssfqrpvsdmniaagcpilfcvskmeaknsyvdafikai 495
 QY 500 QGSSRRHLGDAFKEDPNSSFKKPTGEMNIASGCPVFAQTGLE-NGYIKDITIFIKVI 558
 Db 496 vdlcgl 501
 QY 559 VDTSDL 564
 RESULT 2
 ID TRF1 MOUSE STANDARD; PRT; 409 AA.
 AC P39428;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE TNF RECEPTOR ASSOCIATED FACTOR 1 (TRAF1).
 OS MUS MUSCULUS (MAMMALIA).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 123-135 AND 390-402.
 RM 94349371
 RA ROTHE M., WONG S.C., HENZEL W.J., GOEDEL D.V.;
 RL CELL 78:681-692(1994).
 CC -1- FUNCTION: SIGNAL TRANSDUCER ASSOCIATED WITH THE CYTOPLASMIC DOMAIN
 OF THE 75 KD TUMOR NECROSIS FACTOR RECEPTOR (TNF-R2).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SUBUNIT: HETERODIMER OF TRAF1 AND TRAF2.
 DR EMBL; L33302; MTRAF1A.
 KW COILED COIL.
 SQ SEQUENCE 409 AA; 45464 MM; 857796 CN;

Query Match 19.1%; Score 796; DB 7; Length 409;
 Best Local Similarity 38.8%; Pred. No. 1.22e-122;
 Matches 149; Conservative 94; Mismatches 117; Indels 24; Gaps 18;
 Db 40 lrrdedricpkradn-lhpsvpsplgtg-kh-edvaeaaimcpgavgsfsgspqs 96
 QY 188 LQKHEDTDCP-CVAVSCPHHCQSVOTLLRSLSHLSCEVNAFSTCFKRYGCVFGTNOQ 246
 Db 97 mgeheatsqeshlylllavlkewkspsgnlgaapm-alernuiseiqgaaveatgdlv 155
 QY 247 IKAHDASSAVOHVLL-----KENSNSLEKRVSLQNESVEKRSIOSLINQI-CSEFEI 301
 Db 156 dcyracpccsqee-lalqhlvk-e-kllaqlseeklvf-an-lva-vlnkvea--shla 207
 QY 302 EROKEMLNNSKSLIHLQVVIDSQAEKIKELDKETIRPFQNMEDSMKSVESLQNNRT 361
 Db 208 laaslng--qltrehllslegvvelqgtlaqkdqvlghlsilmeaasfdgtflwk 265
 QY 362 ELESVDKSAAGVARNHGTGL-LESQLSRHDQMLSVHDIRLADMDIGFOVLETAISINGVLWK 420
 Db 266 ltnvrkheavcgrtvalfspafyrtakvgvkiicliylngdsgckthlsfivlmrge 325
 QY 421 IROTKRRKQEVNMGKTI 480
 Db 336 ydallpwpfrnkvtfmldq-nurehaidafpdrsssfqrpvsdmniaagcpilfcvsk 384
 QY 481 YDALLPWFKQKVTMLMDQSSRRHLGDAFKEDPNSSFKKPTGEMNIASGCPVFAQT 540
 Db 385 klgsphayvktdmfcikvdr 408
 QY 541 VLENG--TYIKDITIFIKVIDTS 562
 RESULT 3
 ID DG17 DICD1 STANDARD; PRT; 458 AA.
 AC P11467;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE DG17 PROTEIN.
 GN ZFAA OR DG17.
 OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
 CC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; SARCODINA; RHIZOPODA;
 CC EUMYCETIZOZA; DICTYOSTELIA.
 [1]
 RN SEQUENCE FROM N.A.
 RM 88142840
 RA DRISCOLL D.M., WILLIAMS J.G.;
 RL MOL. CELL. BIOL. 7:4482-4489(1987).
 CC -1- THE EXPRESSION OF DG17 PROTEIN IS DEVELOPMENTALLY REGULATED.
 CC -1- INDUCTION: BY CAMP DURING AGGREGATION.
 DR EMBL; M18106; DDDG17A.
 KW PIR; A29361; A29361.
 DR DICTYOB; DD02010; ZFAA.
 KW DEVELOPMENTAL PROTEIN; ZINC-FINGER.
 FT SIMILAR 25 67 TO DROSOPHILA SINA (AA 71-108).
 FT ZN_FING 27 66 POTENTIAL.
 FT ZN_FING 178 198 POTENTIAL.
 SQ SEQUENCE 458 AA; 53015 MM; 1107348 CN;


```

Query Match Similarity 4.6%; Score 191; DB 2; Length 458;
Best Local Similarity 25.8%; Pred. No. 1.82e-12;
Matches 39; Conservative 36; Mismatches 63; Indels 11; Gaps 10.

Db 124 kdeengckekievdgidshlhn-cqykfvrcsfkyc-ekllr--m-nslkmngfklytc 178
      :: | | : : : | | : : : | | : : : | | : : : | | : : : | |
Qy 118 RNSRRCGAEDLMGHL-VHLKNDCHFELPCVRDCKEKYLRDLBDHYEAKACKYREATNC 176

Db 179 dfctrdikkketeltykr-cmpripdcsgskv-ierksidhndcentqpkcyf 236
      || : : : | : | | | : : | : | | : : | : | | : : | : | :
Qy 177 SHCK-SQVPMIALQKHEIDTCGVVSCPHKCSVQTLLRELSAHL-SECVNAPSTCSFK 234

Db 237 egqckvekmkrkshqnlervnhqymgllie 267
      || : : : | : : : | : : : | : : : | : : : | : : : | : :
Qy 235 RYCGVFGCTNQGIKAHEASSAVQH-VNLKE 264

RESULT 4
AC MEPA MOUSE STANDARD; PRT; 760 AA.
AC P28825;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE MEPRIN A ALPHA-SUBUNIT PRECURSOR (EC 3.4.24.18) (ENDOPEPTIDASE-2)
DS (MEP-1).
OS MOS MUSCULUS (MOUSE).
OC EKKAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=C57BL/6, AND C3H/HE; TISSUE=KIDNEY;
RM 9250517
RA JIANG W., GORBEA C.M., FLANNERY A.V., BEYNON R.J., GRANT G.A.,
RA BOND J.S.;
RN J. BIOL. CHEM. 267:9185-9193(1992).
RP [2]
RP SEQUENCE OF 77-275 FROM N.A.
RM 92042028
RA DOMERNUCH E., STERCHI E.E., JIANG W., WOLZ R.L., BOND J.S.,
RA FLANNERY A.V., BEYNON R.J.;
RN J. BIOL. CHEM. 266:21381-21385(1991).
RP [3]
RP CHARACTERIZATION.
RM 91355206
RA WOLZ R.L., HARRIS R.B., BOND J.S.;
RL BIOCHEMISTRY 30:8488-8493(1991).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEIN AND PEPTIDE SUBSTRATES
CC PREFERENTIALLY ON CARBOXYL SIDE OF HYDROPHOBIC RESIDUES.
CC -1- COFACTOR: ZINC.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SUBUNIT: HOMOTETRAMER OF ALPHA OR BETA SUBUNITS, HETEROTETRAMER
CC OF TWO ALPHA AND TWO BETA SUBUNITS ARE FORMED BY NON-COVALENT
CC ASSOCIATION OF TWO DISULFIDE-LINKED HETERODIMERS; GENETIC FACTORS
CC DETERMINE WHICH OLIGOMER(S) WILL BE FORMED (STRAIN-SPECIFIC).
CC -1- PTM: N-GLYCOSYLATED; AT LEAST 3 OF THE POTENTIAL SITES ARE USED.
CC -1- TISSUE SPECIFICITY: KIDNEY, INTESTINAL BRUSH BORDERS, AND
CC SALIVARY DUCTS.
CC -1- SIMILARITY: HIGH TO OTHER ZINC METALLOPROTEASES. BELONGS TO THE
CC ASTHININ SUBFAMILY.

```

CC	-1-	SIMILARITY: THE PROTEIN INCLUDES 1 EGF-LIKE REPEAT.
DR	EMBL; U74897;	MMEPRIN.A.
DR	EMBL; M62962;	HSPH.
DR	PIR; A40195;	A40195.
DR	PDB; 1IAF;	31-AUG-94.
DR	PROSITE; PS00142;	ZINC_PROTEASE.
DR	PROSITE; P500740;	MAM_
KW	HYDROLASE; METALLOPROTEASE; ZINC; GLYCOPROTEIN; TRANSMEMBRANE; ZINC; SIGNAL; EGF-LIKE DOMAIN; 3D-STRUCTURE.	
FT	SIGNAL	1 33
FT	PROPEP	34 77
FT	CHAIN	78 760
FT	DOMAIN	78 726
FT	TRANSSEM	727 754
FT	DOMAIN	755 760
FT	DOMAIN	778 275
FT	DOMAIN	276 425
FT	REPEAT	685 725
FT	METAL	167 167
FT	ACT SITE	168 168
FT	METAL	171 171
FT	METAL	177 177
FT	DISULFID	688 699
FT	DISULFID	693 708
FT	DISULFID	710 723
FT	CARBOHYD	41 41
FT	CARBOHYD	152 152
FT	CARBOHYD	234 234
FT	CARBOHYD	270 270
FT	CARBOHYD	330 330
FT	CARBOHYD	426 426
FT	CARBOHYD	452 452
FT	CARBOHYD	546 546
FT	CARBOHYD	553 553
FT	CARBOHYD	614 614
FT	CONFLICT	272 272
SEQ	SEQUENCE	760 AA; 85702 MW; 22920193 CN;
<hr/>		
Query Match	Similarity	Score 164; DB 5; Length 760;
Best Local	Similarity	30.0%; Pred. No. 1.96e-08;
Matches	27; Conservative	25; Mismatches 34; Indels 4; Gaps 4;
<hr/>		
Db	447 wvrimnsgqlentvtygdkl-v-sprfynse-gygygvntlypngritnsngllqitfnly	503
Oy	418 iwmkroykrkkoeavwmgktlsyspfttgyfgfrcmcarvylngdmkgcthlslfpviim	477
Db	504 sgndadallewventrgaimtlldgeadtin	533
Oy	478 RGEYDALLPWFKKQVTLA-LMDGSSRRH	506
<hr/>		
RESULT	5	
ID	TPM1 LOCMI	STANDARD; PRJ; 283 AA.
AC	P31816;	
DT	01-JUL-1993 (REL. 26, CREATED)	
DT	01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)	
DT	01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)	
DE	TROPOMIOSIN, MOSCLE.	
OS	LOCUSTA MIGRATORIA (MIGRATORY LOCUST).	


```
Query Match Similarity 3.6%; Score 149; DB 7; Length 283;
      Best Local Similarity 19.0%; Pred.No.2.69e-06;
Matches 20; Conservative 36; Mismatches 46; Indels 3; Gaps 3.
```

RESULT	6	STANDARD;	PRT;	377 AA.
ID	RINI HUMAN			
AC	006587;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)			
DE	RINI1 PROTEIN.			
GN	RINI1.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
CC	EUETHERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RM	93211912			
RA	LOVEING R., HANSON I.M., BORDEN K.L.B., MARTIN S., O'REILLY N.J.,			
RA	EVAN G.T., RAHMAN D., PAPEIN D.J.C., THOMASDALE J., FREEMONT P.S.;			
RL	PROC. NATL. ACAD. SCI. U.S.A. 90:2112-2116(1993).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.			
EMBL	214000; HSRING1.			
DR	PIR; A47380; A47380.			
DR	HSP; P28990; 1CHC.			
DR	PROSITE; P500518; ZINC FINGER C3HC4.			
KW	ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.			
FT	ZN FING 19 58			
FT	DOMAIN 176 231			
FT	DOMAIN 285 348			
FT	DOMAIN 172 175			
SEQUENCE	377 AA; 39145 MW; 716288 CN;			

ID	RESULT	7	STANDARD;	PRT;	1969 AA.
AC	MYSA CAEEL				
AC	P128f4;				
DT	01-OCT-1989	(REL. 12, CREATED)			
DT	01-OCT-1989	(REL. 12, LAST SEQUENCE UPDATE)			
DT	01-JUN-1994	(REL. 29, LAST ANNOTATION UPDATE)			
DE	MYOSIN HEAVY CHAIN A (MHC A).				
GN	MYO-3.				
OS	CERONRABDITIS ELEGANS.				
OC	ERAKRYOTA; METAQOA; ACOELIEMATES; NEMATODA; SECCERENITEA; RHABDITIDA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL NZ;				
RM	89178677				
EA	DIEB N.J., MARDYAMA I.N., KRAUSE M., KARN J.;				

CC	-1-	SUBUNIT: MUSCLE MOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)	
CC	-1-	AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).	
CC	-1-	EACH MOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMOSIN (LM) AND 1 HEAVY MEROMOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).	
CC	-1-	DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.	
CC	-1-	SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.	
CC	-1-	SIMILARITY: THE GLOBULAR HEAD A SEQUENCE SHOWS A HIGH DEGREE OF SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.	
CC	-1-	PTM: TWO CYSTEINE REQUIRED FOR THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MOSIN ATPASE ACTIVITY.	
CC	-1-	THESE ARE FOUR DIFFERENT MOSIN HEAVY CHAINS IN C. ELEGANS.	
CC	-1-	MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.	
DR	ENML;	X08067; CEMYO3.	
DR	PIR;	S02771; S02771.	
DR	HSP7;	P24733; 15CM.	
KW	MOSIN;	MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING; ATP-BINDING; METHYLATION; ALKYLATION; HEPTAD REPEAT PATTERN;	
KW	MLTIGENE	FAMILY.	
FT	DOMAIN	1 851 GLOBULAR HEAD (S1).	
FT	NP BIND	852 1969 RODLIKE TAIL (S2 AND LM DOMAINS).	
FT	DOMAIN	179 186 ATP.	
FT	DOMAIN	667 689 ACTIN-BINDING.	
FT	DOMAIN	770 784 ACTIN-BINDING.	

FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 707 707 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 717 717 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1969 AA; 225509 MW; 1552438 CN;
 Query Match 3.6%; Score 151; DB 5; Length 1969;
 Best Local Similarity 21.2%; Pred. No. 1.41e-06;
 Matches 28; Conservative 49; Mismatches 47; Indels 8; Gaps 7;
 Db 985 dhnrsdqmdaevaklnk-khgesnrklnedqgeekvnhlekinklqeq 1043
 QY 270 EKKVSLQNESVEKRSKISQSLHNOICFSEIEIRQ-KEMRNRESKILHLQWVIDSQAEK 328
 Db 1044 mdeleandrekeregkalekkrkveq-dlkvaq-enideatekq--kh-dv-etlkrke 1097
 QY 329 IKELDKREIREFPRQWNEADSMKSSVESLQNRVTELESVDKSGQVARNIGLESQLSRHD 388
 Db 1098 edlhtnklae 1109
 QY 389 QMLSVHDRLAD 400
 RESULT 8
 ID MYSN CHICK STANDARD; PRT; 1959 AA.
 AC P14105;
 DT 01-JUN-1990 (REL. 13, CREATED)
 DT 01-JUN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, NONMUSCLE (NMHC).
 OS GALLUS GALLUS (CHICKEN).
 OC EDUAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 CC GALLIFORMES.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=INTESTINAL EPITHELIUM;
 RM 90046668
 RA SHOHET R.V., CONTI M.A., KAWAMOTO S., PRESTON Y.A., BRILL D.A.,
 RL ADLSTEIN R.S.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:7726-7730(1989).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CAPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)
 AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATTER BE SPLIT FURTHER INTO
 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST
 TO OTHER NONMUSCLE MYOSINS.
 DR EMBL; M26510; GGMVHN.
 DR PIR; A33977; A33977.
 DR HSSP; P24733; 1SCM.
 KW MYOSIN; COILED COIL; ACTIN-BINDING; ALKYLATION; ATP-BINDING;
 HM HEPTAD REPEAT PATTERN; MULTIGENE FAMILY.
 FT DOMAIN 1 835 GLOBULAR HEAD (S1).
 FT DOMAIN 836 1926 RODLIKE TAIL (S2 AND LM DOMAINS).

FT NP BIND 174 181 ATP.
 FT DOMAIN 654 676 ACTIN-BINDING.
 FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1959 AA; 226502 MW; 15394260 CN;
 Query Match 3.5%; Score 145; DB 5; Length 1959;
 Best Local Similarity 25.2%; Pred. No. 9.59e-06;
 Matches 34; Conservative 34; Mismatches 57; Indels 10; Gaps 8;
 Db 1460 yaeedtraeaaareketakalslaleaelegkaelerunkqtr-t--emedlmskcdv 1516
 QY 297 FEILIERQKEMLRNNEKILHLQWVIDSQAEKIKELDKREIREFPRQWNEADSMKSSVESL 356
 Db 1517 gksvhelekakraleqveemktql-e-el--edelqatadaklr-levngam-kagidr 1571
 QY 357 QNRVTELESVDKS-AGQVARNIGLESQLSRHDQMLSVHDRLADMDLGFVLETRSYN- 414
 Db 1572 dligrdeqneekrkq 1586
 QY 415 GVLWKIRDYKRRQ 429
 RESULT 9
 ID MYSN HUMAN STANDARD; PRT; 1961 AA.
 AC P35579;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, NONMUSCLE TYPE A (CELLULAR MYOSIN HEAVY CHAIN,
 DE TYPE A) (NMHC-A).
 OS MYH9.
 OS HOMO SAPIENS (HUMAN).
 OC EDUAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 CC [1]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RM 92003925
 RA TOOTHAKER L.E., GONZALEZ D.A., TUNG N., LEMONS R.S., LE BEAU M.M.,
 RA ARNAOUT M.A., CLAYTON L.K., TENEN D.G.;
 RL BLOOD 78:1826-1833(1991).
 RL [2]
 RP SEQUENCE OF 1-715 FROM N.A.
 RM 91316803
 RA SIMONS M., WANG M., MCBRIDE O.W., KAWAMOTO S., YAMAKAWA K.,
 RA GDULA D., ADLSTEIN R.S., WEIR L.;
 RL CIRC. RES. 69:530-539(1991).
 RL [3]
 RP SEQUENCE OF 715-1961 FROM N.A.
 RM 90138958
 RA SAEZ C.G., MYERS J.C., SHOWS T.B., LEINWAND L.A.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:1164-1168(1990).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CAPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)
 AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATTER BE SPLIT FURTHER INTO

CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST
CC TO OTHER NONMUSCLE MYOSINS.
DR EMBL; M81105; HSMYH9.
DR EMBL; M69180; HSMYOHCA.
DR EMBL; M31013; HSMYONM.
DR HSSP; P24733; 1SCM.
DR MIM; 160775; 11TH EDITION.
DR MYOSIN; COILED COIL; ACTIN-BINDING; ALKYLATION; ATP-BINDING;
KW HEPTAD REPEAT PATTERN; MULTIGENE FAMILY.
FT DOMAIN 1 835 GLOBULAR HEAD (S1).
FT NP BIND 836 1926 RODLIKE TAIL (S2 AND IMM DOMAINS).
FT DOMAIN 174 181 ATP.
FT MOD BIND 654 676 ACTIN-BINDING.
FT MOD RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 53 55 EAI -> HGH (IN REF. 2).
FT CONFLICT 660 660 T -> S (IN REF. 2).
FT CONFLICT 869 869 T -> M (IN REF. 3).
FT CONFLICT 931 931 C -> Y (IN REF. 3).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 3).
SQ SEQUENCE 1961 AA; 226600 MW; 15612344 CN;

Query Match 3.4%; Score 141; DB 5; Length 1961;
Best Local Similarity 27.8%; Pred. No. 3.37e-05;
Matches 32; Conservative 26; Mismatches 48; Indels 9; Gaps 7;

Db 1461 yaeerdaaeareteketakalalarleameqkaelerlnkqfr-t--emedlmskadv 1517
Qy 297 FEIERHOKEMLRNNEKSLILHLOVIDSQAEKKEIKERPRQWWEADSKSSVSL 356
Db 1518 gksvshlekakralaqyeemkqtleelqatdaklrl-evnlgamkqfe 1571
Qy 357 QNRVTELESVDKS-AGQVA-RNTGL--LESQI-SRHDQMLSVHDIRLADMIDIGFQ 406

RESULT 10
ID MYSB CAEEL STANDARD; PRT; 1966 AA.
AC P02566;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN B (MHC B).
CN UNC-54 OR MTO-4.
OS CAENORHABDITIS ELEGANS.
OC EDKARYOTA; METAZOA; ACCELEMATES; NEMATODA; SECCERNENTEA; RHABDITIDA.
RN [1]
RN SEQUENCE FROM N.A.
RM 8373600
RA KARN J., BRENNER S., BARNETT L.;
RL PROC. NATL. ACAD. SCI. U.S.A. 80:4253-4257 (1983).
RN [2]
RN SEQUENCE OF 850-1966 FROM N.A.
RP 82272395
RA MCILACHLAN A.D., KARN J.;
RL NATURE 299:226-231 (1982).
RN [3]

RP SEQUENCE OF 1876-1966 FROM N.A.
RM 8332892
RA WILLS N., GESTELAND R.F., KARN J., BARNETT L., BOLLEN S.,
RA WATERSTON R.H.;
RL CELL 33:575-583 (1983).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)
CC AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATTER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
CC SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
CC HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
CC THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
CC THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C. ELEGANS.
CC -1- MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY WALL MUSCLE.
CC THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
DR EMBL; J01050; CEMYONC.
DR EMBL; V01494; CEMY01.
DR PIR; A02992; MMYK.
DR HSSP; P02649; 1LE4.
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
KW ATP-BINDING; METHYLATION; ALKYLATION; HEPTAD REPEAT PATTERN;
KW MULTIGENE FAMILY.
FT DOMAIN 1 849 GLOBULAR HEAD (S1).
FT DOMAIN 850 1966 RODLIKE TAIL (S2 AND IMM DOMAINS).
FT DOMAIN 1165 1166 ALPHA-HELICAL TRILPECE (S2).
FT NP BIND 177 184 ATP (BY SIMILARITY).
FT DOMAIN 665 687 ACTIN-BINDING.
FT DOMAIN 769 783 METHYLATION (TR1-) (POTENTIAL).
FT MOD RES 128 128 ALKYLATION (SH-1).
FT MOD RES 705 705 ALKYLATION (SH-2).
FT MOD RES 715 715 HINGE.
FT DOMAIN 1165 1176 E -> R (IN REF. 2).
FT CONFLICT 1337 1337 I -> L (IN REF. 2).
FT CONFLICT 1880 1880 I -> L (IN REF. 2).
SQ SEQUENCE 1966 AA; 225125 MW; 15251009 CN;

Query Match 3.4%; Score 142; DB 5; Length 1966;
Best Local Similarity 24.2%; Pred. No. 2.46e-05;
Matches 32; Conservative 37; Mismatches 55; Indels 8; Gaps 8;

Db 984 dbrqrsldemqgqdaaklnkek-khgeelnrkmedtqsecdgmbnkvkqlgt 1042
Qy 270 EKKVSLLONESEVERKNSISLOHQCISFELEIERQ-KEMLRNNEKSLILHLOVIDSQAEK 328
Db 1043 lddledslerakradldkqtkrvqg-elkiag-enides-gr-qth-dl-ennlkke 1096
Qy 329 LKEIDKEIRPRQWWEADSKSSVSESLQNRVTELESVDKSAGQVARNITGLLESQLSRHD 388

ID	RESULT	11	STANDARD;	PRT;	1978 AA.
AC	MYSG CHICK				
AC	P10587;				
DT	01-JUL-1989	(REL. 11,	CREATED)		
DT	01-JAN-1990	(REL. 13,	LAST SEQUENCE UPDATE)		
DT	01-JUN-1994	(REL. 29,	LAST ANNOTATION UPDATE)		
DE	MYOSIN HEAVY CHAIN,	GIZZARD	SMOOTH MUSCLE.		
OS	GALLUS GALLUS (CHICKEN).				
OC	BEKARAYOTA, METRIZOA,	CHORDATA,	VERTEBRATA,	TETRAPODA,	AVES;
OC	GALLIFORMES.				
RN					
RN	[1]				
RN	SEQUENCE FROM N.A.				
RA	88118918				
RA	YANAGISAWA M.,	HAMADA Y.,	KATSURAGAWA Y.,	IMAMURA M.,	MIKAWA T.,
RA	MASAKI T.,				
RL	J. MOL. BIOL.	198:143-157	(1987).		
RP					
RP	REVISIONS.				
RA	MASAKI T.;				
RL	SUBMITTED (FEB-1989)	TO EMBL/GENBANK/DBJ	DATA BANKS.		
RP	[3]				
RP	SEQUENCE OF 1-203.				
RA	88032919				
RA	MAITA T.,	ONISHI H.,	YAJIMA E.,	MATSUDA G.;	
RL	J. BIOCHEM.	102:133-145	(1987).		
CC	-1- FUNCTION: MUSCLE CONTRACTION.				
CC	-1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2				
CC	HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)				
CC	AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).				
CC	-1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMH)				
CC	AND 1 HEAVY MEROMYOSIN (HMH). IT CAN LATER BE SPLIT FURTHER INTO				
CC	2 GLOBAL SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).				
CC	-1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING				
CC	CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.				
CC	CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.				
CC	-1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.				
CC	-1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,				
CC	WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE				
CC	CONSERVED.				
CC	-1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY				
CC	ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.				
DR	EMBL; X06346; GEMVHGSW.				
DR	PIR; S03166; S03166.				
DR	HSSP; P02259; IGHC.				
KW	MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;				
KW	APP-BINDING; METHYLATION; ALKYLATION; HEPTAD REPEAT PATTERN;				
KW	MULTIGENE FAMILY.				
FT	INIT MET	0			
FT	MOD RES	1			
FT	DOMAIN	1	848	BLOCKED.	
FT	DOMAIN	849	1978	GLOBULAR HEAD (S1).	
FT	NP BIND	176	183	RODLIKE TAIL (S2 AND LMH DOMAINS).	
FT	DOMAIN	666	688	APP.	
FT	DOMAIN	767	781	ACTIN-BINDING.	
FT	MOD RES	127	127	ACTIN-BINDING.	
FT				METHYLATION (TRI-) (POTENTIAL).	

[illegible]

ID	RESULT	12	STANDARD;	PRT;	245 AA.
AC	TEMPZ RAT				
DT	P1834d;				
DT	01-NOV-1990 (REL. 16, CREATED)				
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)				
DT	01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)				
DE	TROPOMYOSIN ALPHA CHAIN, BRAIN-3 (TMBR-3).				
GN	ALPHA-TM.				
OS	RATUUS NORVEGICUS (RAT).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	OTHERIA; RODENTIA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRAIN;				
RM	90205854				
RA	TEES-MILLER J.P., GOODWIN L.O., HELFMAN D.M.;				
RL	MOL. CELL. BIOL. 10:1723-1742(1990).				
CC	-1- FUNCTION: THE FUNCTION OF TROPOMYOSIN IN SMOOTH MUSCLE AND NON-				
CC	MUSCLE CELLS IS NOT CLEAR.				
CC	-1- SUBUNIT: DIMER OF AN ALPHA AND A BETA CHAIN.				
CC	-1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE				
CC	EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.				
CC	-1- TISSUE SPECIFICITY: BRAIN				
CC	-1- ALTERNATIVE PRODUCTS: THE TROPOMYOSIN ALPHA-TM GENE PRODUCES, BY				
CC	ALTERNATIVE SPLICING, THE FOLLOWING PROTEINS: STRIATED MUSCLE TM,				
CC	SMOOTH MUSCLE TM, BRAIN TMBR-1, TMBR-2, TMBR-3, FIBROBLAST TM-2,				
CC	TM-3, TM-5A, AND TM-5B.				
DR	EMBL; M34136; RNTMBR3A.				
DR	PIR; C34787; C34787.				
DR	PROSITE; P500326; TROPOMYOSIN.				
DR	BRAIN COILED COIL; REPEAT; ALTERNATIVE SPLICING; MULTIGENE FAMILY.				
SO	SEQUENCE 245 AA; 28343 MW; 256960 CN;				
DB	Query Match	3.1%;	Score 128;	DB 7;	Length 245;
	Neut Local Similarity 23.5%;		Pred. No. 1.75e-03;		
	Matches 20; Conservative 28;	Mismatches 34;	Indels 3;	Gaps 3;	
	31 lqreld-qerk retaadvaa nriq vee ldqrer atqlk leaeakadeeer 89				

QY 318 LQVRIDSQAELKEL-DKEIRPF-RQWWEADSKSSVSLQNRVTELESVDKSGQVAR 375
Db 90 gmkvleaqdeekmeigaiqlke 114
QY 376 NTGLESLSRHDMQMLSVHDIRLAD 400

RESULT 13
ID TPRM RAT STANDARD; PRT; 251 AA.
AC P18343;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE TROPOMYOSIN ALPHA CHAIN, BRAIN-2 (TMR-2).
GN ALPHA-TM.
OS RATTUS NORVEGICUS (RAT).
OC EURAROTI; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RM 90205854
RA LEES-MILLER J.P., GOODWIN L.O., HELFMAN D.M.;
RL MOL. CELL. BIOL. 10:1729-1742(1990).
CC -1- FUNCTION: THE FUNCTION OF TROPOMYOSIN IN SMOOTH MUSCLE AND NON-MUSCLE CELLS IS NOT CLEAR.
CC -1- SUBUNIT: DIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- ALTERNATIVE PRODUCTS: THE TROPOMYOSIN ALPHA-TM GENE PRODUCES, BY ALTERNATIVE SPLICING, THE FOLLOWING PROTEINS: STRIATED MUSCLE TM, SMOOTH MUSCLE TM, BRAIN TMR-1, TMR-2, TMR-3, FIBROBLAST TM-2, TM-3, TM-5A, AND TM-5B.
DR EMBL; M34134; RUTHER2A.
DR PIR; B34787; B34787.
DR PROSITE; P500326; TROPOMYOSIN.
KW BRAIN; COILED COIL; REPEAT; ALTERNATIVE SPLICING; MULTIGENE FAMILY.
SQ SEQUENCE 251 AA; 28704 MW; 298473 CN;

Query Match 3.1%; Score 128; DB 7; Length 251;
Best Local Similarity 23.3%; Pred. No. 1.75e-03;
Matches 20; Conservative 28; Mismatches 34; Indels 3; Gaps 3;
Db 31 lgrld-gerklrtlaadvaslnrtqlveeldragerlatalkleaeakaadeer 89
QY 318 LQVRIDSQAELKEL-DKEIRPF-RQWWEADSKSSVSLQNRVTELESVDKSGQVAR 375
Db 90 gmkvleaqdeekmeigaiqlke 114
QY 376 NTGLESLSRHDMQMLSVHDIRLAD 400

RESULT 14
ID TPRM RAT STANDARD; PRT; 257 AA.
AC P04462;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EURAROTI; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RM 85030490
RA PERIASAMY M., WIECZOREK D.F., NADAL-GINARD B.;
RL J. BIOL. CHEM. 259:13573-13578(1984).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC), AND 2 REGULATORY LIGHT CHAIN SUBUNITS (RLC-2).
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEMOROSIN (LM) AND 1 HEAVY MEMOROSIN (HM). IT CAN LATTER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES. CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SOBELDULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED DURING PERINATAL SKELETAL MUSCLE DEVELOPMENT. IT IS FIRST DETECTED LATE IN FETAL LIFE, IS MAXIMALLY EXPRESSED AT THE END OF THE 1ST POSTNATAL WEEK, AND IS NOT FOUND IN THE ADULT. ALTHOUGH THEIR EXPRESSION TEMPORALLY OVERLAPS.
CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES, WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC -1- THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN THE RAT.
DR EMBL; K02111; RMYHC.
DR PIR; A02991; A02991.
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING; ATP-BINDING; HEPTAD REPEAT PATTERN; MULTIGENE FAMILY.
FT NON TER 1 1
FT DOMAIN <1 257 LIGHT MEMOROSIN (LM).
SQ SEQUENCE 257 AA; 29843 MW; 288053 CN;

Query Match 3.1%; Score 131; DB 5; Length 257;
Best Local Similarity 22.3%; Pred. No. 7.18e-04;
Matches 39; Conservative 45; Mismatches 83; Indels 8; Gaps 8;
Db 42 gnaelina-kkklendvsgqsevevigrarnaekakkaitdaamaealkqcdsa 100
QY 257 QHVNLIKEMWSLEKVSLLQNSVEKNKSISLIHQICSPFEIEROKEMIRNNESKITL 316
Db 101 hlermkmevtvdkldhrde-aeglaikgokq-ickleatveleg-eveneg-krn 156
QY 317 HLQVRIDSQAELKEL-DKEIRPF-RQWWEADSKSSVSLQNRVTELESVDKSGQVAR 376
Db 157 aavkv-lrkheerikelvg-teedrk-nvlrlgdlvklgakyksykrgeea 208
QY 377 TGLESLSRHDMQMLSVHDIRLADMDLGFVLETAYNGVLLMKIRDKRRRQEA 431

RESULT 15
ID TPRM TRICO STANDARD; PRT; 284 AA.
AC P15846;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)

ALIGNMENTS

```

RESULT 1
ID R5148 standard; Protein; 475 AA.
AC R5148;
DT 14-FEB-1992 (first entry)
DE Ro/SSA autoantigen.
KW Autoantibody; autoantigen; SLE; systemic lupus erythematosus.
OS Homo sapiens.
PN M09117171-A.
PD 14-NOV-1991.
PF 07-MAY-1991; 003139.
PR 07-MAY-1990; DS-520270.
RA (OKLA-) OKLAHOMA MED RES FO.
PI Frank MB, Itoh K.
DR WPI; 91-353712/48.
DR N-PSDB; Q14798.
PT DNA encoding an Ro-SSA autoantigen - useful for diagnosing
PT auto-immune disorders or presence of auto-antibodies
PS Disclousure; Fig 2; 41p; English.
CC A cDNA library (from human thymus mRNA) in lambda gtl1 was screened
CC with serum from a patient having systemic lupus erythematosus. Two
CC clones were reactive with sera (from a panel of lupus patients)
CC which contd. autoantibodies against 52 kD protein.
CC Both the cDNA and the protein expressed from it, or portions of it,
CC are useful as diagnostic agents in the identification of patients
CC having autoantibodies and in the identification and analysis of
CC the structural and functional properties of the autoantigen and for
CC application in immunotherapeutic regimens.
SQ Sequence 475 AA;
Query Match 28.4%; Score 107; DB 3; Length 475;
Best Local Similarity 27.3%; Pred. No. 1.08e-02;

```

	Matches	12; Conservative	13; Mismatches	19; Indels	0; Gaps	0; Ns
Db	16	cpicldpfvpevysiecgnsfcgeciqsqykggsvcpvcrqrfl	59			
		: : : : : : : : :				
Qy	5	CECRHVLVLCSPKQTECGHRFCESGMALILSSSPKCTACGEIV	48			

ID	RESIDUE	SEQUENCE
AC	15-MAY-1995	(first entry)
DE	Herpes simplex virus ICP0 protein domain homologous to part of EPO.	
KM	Pseudotabular virus; PRV; LTV; large latency transcript;	
KW	attenuated virus; vaccine; early protein 0; EPO; HSV-1 ICP0;	
KW	protecting animals; deletion mutants; swine; cysteine-rich;	
KW	zinc finger motif.	
OS	Herpes simplex virus 1.	
FT	Key	Location/Qualifiers
FT	Misc.difference 6	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
FT	Misc.difference 9	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
FT	Misc.difference 27	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
FT	Misc.difference 30	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
FT	Misc.difference 41	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
FT	Misc.difference 44	
FT	/note= "cysteine residue that is part of the zinc	
FT	finger motif"	
PN	US3352596-A.	
PD	04-OCT-1994.	
PR	11-SEP-1992; 945283.	
PR	11-SEP-1992; US-945283.	
PA	(USDA) US SEC OF AGRIC.	
PI	Cheung AK, Wesley RD;	
PI	WPI; 94-316187/39.	
PT	New pseudotabular virus mutants for use in vaccine - having a	
PT	deletion and/or insertion in the early protein 0 gene or large	
PT	latency transcript gene	
PS	disclosure; Column 45-47; 43pp; English.	
CC	R60622-24 show cysteine-rich zinc finger domains that is conserved among	
CC	herpesviruses. This peptide is the homologous region from the Herpes	
CC	simplex virus 1. This is the most important domain of the EPO gene. The	
CC	DNA encoding this cysteine-rich domain (see 073501) also encodes an amino	
CC	acid sequence specific to the LIT gene in the opposite orientation.. Thus	
CC	deletion in the EPO gene other than the first 200 bp of EPO will	
CC	automatically delete some of the DNA sequences encoding the LIT. EPO is	
CC	nonessential for replication, LIT (large latency transcript) is the only	
CC	gene expressed during PRV latency, and the IE180 gene is absolutely	
CC	necessary for PRV replication. However there are 2 copies of IE180 in the	
CC	genome. It is expected that PRV lacking one of the IE180 copies is	
CC	viable. Deletions in the non-overlapping regions of these 3 genes will	
CC	generate single deletion mutants, while deletions in overlapping regions	

CC will generate double deletion mutants. The invention is concerned with
 CC the construction of attenuated viruses which have a reduced ability to
 CC reactivate from latency. This can be achieved by functionally disabling
 CC the expression of the EPO gene, or by disrupting the synthesis of the
 CC LIT, or both. (See also Q13500-1 and R60620-24)
 SQ Sequence 69 AA;

Query Match 22.0%; Score 83; DB 11; Length 69;
 Best Local Similarity 34.5%; Pred. No. 1.49e+00;
 Matches 10; Conservative 5; Mismatches 13; Indels 1; Gaps 1;

Db 24 cmhfcipckmktmqlnt-cplonaklv 51
 QY 20 CGHRFCSCMAALSSSPKTCACQESIV 48

RESULT 3

ID P91461 standard; peptide; 365 AA.

AC P91461;

DT 10-MAR-1993 (revised)

DT 09-JAN-1990 (first entry)

DE Rpt-1 clone cDNA insert.

KW Regulatory protein T lymphocyte-1; IL-2 receptor;

KW human lymphotropic retroviruses; metal fingers.

OS Homo sapiens.

FN W08907652-A.

PD 24-AUG-1989.

PT 10-FEB-1989; U00550.

PR 11-FEB-1988; US-154758.

PA (DANA) Dana-Farber Cancer Institute.

PI Cantor HI, Patarca R, Freeman GJ;

DR WPI; 89-263719/36.

PT Regulatory protein T lymphocyte-1 gene and corresp. protein - increase

PT gene expression of interleukin-2 receptor and of human lymphotropic

PT retroviruses.

PS Claim 20; Fig 2; 58pp; English.

CC cDNA was prep'd. from L cells (fibroblast tumour) and 2PK3 (B cell

CC lymphoma). Host was pref. E. coli NRRL B-18297. Cysteine residues at

CC positions 15, 18, 35, 38, 55, 58, 115 and 118, and histidine residues at

CC positions 124 and 127 may be involved in metal finger formation.

CC Positions are given from the methionine (a.a 13 in the sequence given).

CC See also N90702.

CC Sequence 365 AA;

SQ Sequence 365 AA;

Query Match 21.8%; Score 82; DB 1; Length 365;

Best Local Similarity 34.6%; Pred. No. 1.82e+00;

Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Db 27 cpiclellkpsadchsfrcacit 52

QY 5 CECHIVLCSPKQTECGHRFCSCMA 30

RESULT 4

ID R27535 standard; Protein; 560 AA.

AC R27535;

DT 09-MAR-1993 (first entry)

DE myl protein.

KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;

KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;

KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.

OS Synthetic.

PN W09216660-A.

PD 01-OCT-1992.

PF 23-MAR-1992; U02320.

PR 22-MAR-1991; US-673838.

PR 22-MAR-1991; US-675084.

PA (SLOK) SLOAN KETTERING INST CANCER.

PI Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;

PI Warrell RP;

DR WPI; 92-349240/42.

DR N-PSDB; Q29354.

PT Marker for acute promyelocytic leukaemia and other neoplasias -

PT comprising nucleic acid and encoded abnormal retinoic acid

PT receptor-alpha receptor

PS Disclosure; Page 48-51; 84pp; English.

CC The sequence given represents the amino acid sequence of one variation

CC of the myl protein. The gene encoding this protein is disrupted in a

CC translocation of a portion of the long arm of chromosome 17 onto the

CC long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This causes a fusion

CC between retinoic acid receptor (RAR)-alpha and myl which is

CC characteristic of acute promyelocytic leukemia (APL). The breakpoint

CC region has been cloned and it has been shown that DNA rearrangements

CC are clustered in the region of the first intron of RAR-alpha. This

CC sequence was isolated by polymerase chain reaction (PCR). The primers

CC used for amplification of this sequence can also be used to amplify the

CC translocated region.

CC Sequence 560 AA;

SQ Sequence 560 AA;

Query Match 20.2%; Score 76; DB 5; Length 560;

Best Local Similarity 34.5%; Pred. No. 5.90e+00;

Matches 10; Conservative 6; Mismatches 11; Indels 2; Gaps 2;

Db 56 rcgqgaekc-pklpchlhcagclea 83

QY 4 KCERCHL-VLCSPKQTECGHRFCSCMA 31

RESULT 5

ID R27533 standard; Protein; 797 AA.

AC R27533;

DT 09-MAR-1993 (first entry)

DE myl/RAR-alpha fusion.

KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;

KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;

KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.

OS Synthetic.

PN W09216660-A.

PD 01-OCT-1992.

PF 23-MAR-1992; U02320.

PR 22-MAR-1991; US-673838.

PR 22-MAR-1991; US-675084.

PA (SLOK) SLOAN KETTERING INST CANCER.

PI Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;

PI Warrell RP;

DR WPI; 92-349240/42.

DR N-PSDB; Q29334.

PT Marker for acute promyelocytic leukaemia and other neoplasias -

PT comprising nucleic acid and encoded abnormal retinoic acid

PT receptor-alpha receptor

PS Disclosure; Fig 4; 84pp; English.
 CC The sequence given shows a fusion between retinoic acid receptor
 CC (RAR)-alpha and myl which is characteristic of acute promyelocytic
 CC leukemia (APL). This is caused by a translocation of a portion of
 CC the long arm of chromosome 17 onto the long arm of chromosome 15
 CC [t(15;17)(q21;q11-22)]. The breakpoint region has been cloned and it
 CC has been shown that DNA rearrangements are clustered in the region
 CC of the first intron of RAR-alpha. This sequence was isolated by
 CC polymerase chain reaction (PCR) using primers which correspond to
 CC sequences both 5' and 3' to the breakpoint region.
 SQ Sequence 797 AA;

Query Match 20.2%; Score 76; DB 5; Length 797;
 Best Local Similarity 34.5%; Pred. No. 5.90e+00;
 Matches 10; Conservative 6; Mismatches 11; Indels 2; Gaps 2;

Db 56 rcgqgaekc-pkllpchlntscgclea 83
 Qy 4 KCEKCHL-VLCSPKTECGHRCSCMAA 31

RESULT 6

ID RI0533 standard; Protein; 165 AA.
 AC RI0533;
 DT 12-APR-1991 (first entry)
 DE Prod. of pMG4B12 used to isolate style-stigma specific gene STG4B12.
 KW Hybrid vigour; pollination.
 OS Nictiana tabacum "Petit Havana" SRI.
 PN EP-412006-A.
 PD 06-FEB-1991.
 PF 31-JUL-1990; 402196.
 PR 04-AUG-1989; EP-402224.
 PR 31-JUL-1990; EP-402196.
 PA (PLAN-) PLANT GENETIC SYST.
 PI De Greef W, Van Emmelo J, De Oliveira DE, De Souza MH;
 PI Van Montagu M;
 DR N-PSDB; 010376.
 DR WPI; 91-038878/06.
 PT Modified plant with transformed nuclear genome - obtd. by using
 PT foreign DNA encoding prod. which disrupts metabolism, functioning
 PT and/or development of plant.
 PS Disclosure; Fig 2b; 42pp; English.
 CC The clone was produced by subcloning a style-stigma specific cDNA,
 CC 4B12, obtd. from Prof. Goldberg of UCLA, into pCEM1, to produce
 CC pMG4B12. The clone was found to be stigma-style specific in
 CC Northern analysis. A probe from the clone was used to isolate
 CC the corresp. genomic sequence which is specifically expressed in
 CC style-stigma tissues of the female organ of tobacco. The corresp.
 CC clone, pSTG4B12, contains STM4B12, an "STM-type" gene. This and
 CC other similar genes can be used to produce new hybrid plants or
 CC seeds having a combination of desirable traits, and showing hybrid
 CC vigour. A female-sterile, male-fertile plant can be produced which
 CC favours cross-pollination.
 CC See also RI0531-35.
 SQ Sequence 165 AA;

Query Match 19.6%; Score 74; DB 2; Length 165;
 Best Local Similarity 47.8%; Pred. No. 8.69e+00;
 Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 2;

Db 143 ncgkcnv-cspgqk-csfqcd 163
 Qy 4 KCEKCHLVCSPKTECGHRCFCE 26

RESULT 7

ID R60622 standard; Protein; 67 AA.
 AC R60622;
 DT 15-MAY-1995 (first entry)
 DE Pseudorabies virus early polypeptide 0 homologous protein domain.
 KW Pseudorabies virus; PRV; LIV; large latency transcript;
 KW attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICP0;
 KW protecting animals; deletion mutants; swine; cysteine-rich;
 KW zinc finger motif.
 OS Pseudorabies virus.
 FH Key Location/Qualifiers
 FT Misc difference 6
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT Misc difference 9
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT Misc difference 27
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT Misc difference 30
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT Misc difference 41
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT Misc difference 44
 FT /note= "cysteine residue that is part of the zinc
 FT finger motif"
 FT US352596-A.
 PN 04-OCT-1994.
 PD 11-SEP-1992; 945283.
 PF 11-SEP-1992; US-945283.
 PR (USDA) US SEC OF AGRIC.
 PA Cheung AK, Wesley RD;
 PI WPI; 94-316187/39.
 DR New pseudorabies virus mutants for use in vaccine - having a
 PT deletion and/or insertion in the early protein 0 gene or large
 PT latency transcript gene
 PS Disclosure; Column 45-46; 43pp; English.
 CC R60622-24 show cysteine-rich zinc finger domains that is conserved among
 CC herpesviruses. This is the most important domain of the EP0 gene. The DNA
 CC encoding this cysteine-rich domain (see Q73501) also encodes an amino
 CC acid sequence specific to the LIT gene in the opposite orientation. Thus
 CC deletion in the EP0 gene other than the first 200 bp of EP0 will
 CC automatically delete some of the DNA sequences encoding the LIT. EP0 is
 CC nonessential for replication, LIT (large latency transcript) is the only
 CC gene expressed during PRV latency, and the IE180 gene is absolutely
 CC necessary for PRV replication. However there are 2 copies of IE180 in the
 CC genome. It is expected that PRV lacking one of the IE180 copies is
 CC viable. Deletions in the non-overlapping regions of these 3 genes will
 CC generate single deletion mutants, while deletions in overlapping regions
 CC will generate double deletion mutants. The invention is concerned with
 CC the construction of attenuated viruses which have a reduced ability to
 CC reactivate from latency. This can be achieved by functionally disabling

Matches 9; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Db 46 prllcgtchdcitrlphgrairc 72
1: ||| | | : : : :
Qy 15 PKOTCGHRCESCMAL-LSSSSPKC 40

RESULT 10
ID R66034 standard; Protein; 574 AA.
AC R66034;
DT 05-JUN-1995 (first entry)
DE Rat ARD 1.
KW ADP-ribosylation factor; ARF; ARD 1; probe.
OS Rattus rattus.
PN W09424283-A.
PD 27-OCT-1994.
PT 15-APR-1994; U064190.
PR 16-APR-1993; US-048252.
PS 19-APR-1993; US-049473.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Mishima K, Mose J, Nightingale M, Tsuchiya M;
DR WPI; 94-341862/42.
N-PSDB: Q79327.
PT GTP-binding protein ARD1 with ADP-ribosylation factor domain -
FT useful as biochemical and diagnostic reagent
PS Example; Table 1, Page 11-12; 52pp; English.
CC A novel ARD 1 protein includes an 18 kDa region that
exhibits significant homology to known ADP-ribosylation factors
(ARFs) and is called ARD 1 for ARF domain. Following transfection of
reverse transcribed, PCR amplified human sequences into competent DH5
-alpha cells, clones corresp. to ARD 1 were isolated. The putative ORF
of this gene, termed ARD 1, consisted of 1722 nts encoding a
protein of 574 AAs with an ARF related domain at the carboxy1
terminus (see Q79326/R66033). Clones corresp. to rat ARD 1 were
isolated. A rat brain Lambda ZAP11 library was screened with oligo
CC R26CR. It yielded clone 2a conng. an insert that corresp. to nts
61-1973 of human ARD 1. Q79327 and R66034 give the cDNA and deduced
CC AA sequence of rat ARD1 which corresp. to the ORF of human ARD 1.
CC NT and deduced AA sequences of ARD 1 coding regions from rat and
CC human are 92% and 98% identical, respectively, without any gaps.
SQ Sequence 574 AA;

Query Match 18.8%; Score 71; DB 12; Length 574;
Best Local Similarity 36.4%; Pred. No. 1.54e+01;
Matches 8; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

Db 51 cghtvchdcitrlphgrairc 72
||| | | : : : :
Qy 20 CGHRCESCMAL-LSSSSPKC 40

RESULT 11
ID R63059 standard; Protein; 124 AA.
AC R63059;
DT 15-AUG-1995 (first entry)
DE Rat PLA2 type I.
KW RPLA2-8; RPLA2-10; phospholipase A2; PLA2.
OS Homo sapiens.
PN W09502328-A.
PD 26-JAN-1995.

PF 15-JUL-1994; U07926.
PR 15-JUL-1993; US-091941.
PS 26-JUL-1993; US-097354.
PA (INCYTE-) INCYTE PHARM INC.
DR (INDV) UNIV INDIANA FOUND.
PI Seilhamer JJ, Tischfield JA;
DR WPI; 95-067096/09.
PT Novel type III and IV low mol. wt. phospholipase A2 enzymes -
PT from humans and rats, also nucleic acid sequences useful, e.g.
PT for recombinant prodn. of enzymes, research into Batten's
PT disease, etc.
PS Disclosure; Page 27; 160pp; English.
CC Rat cDNAs (081136-37) encoding 2 novel PLA2s, RPLA2-8 (R63044)
CC and RPLA2-10 (R63045), were isolated from brain and heart
CC libraries, respectively. RPLA2-8 and RPLA2-10 were characterized
CC as novel type III and IV PLA2s on the basis of their cysteine
CC content compared with RPLA2 types I (R63059) and II (R63059).
SQ Sequence 124 AA;

Query Match 18.3%; Score 69; DB 12; Length 124;
Best Local Similarity 34.6%; Pred. No. 2.26e+01;
Matches 9; Conservative 6; Mismatches 9; Indels 2; Gaps 2;

Db 75 ykcsag-ivntcdkndceeficn-c 98
||| : : | : : : :
Qy 3 YKCEKHLVLCSPKOTCGHRCESC 28

RESULT 12
ID R63052 standard; Protein; 146 AA.
AC R63052;
DT 15-AUG-1995 (first entry)
DE HPLA2-8.
KW HPLA2-8; phospholipase A2; PLA2; Batten disease;
KW neuronal ceroid lipofuscinosis; gene therapy.
OS Homo sapiens.
PN W09502328-A.
PD 26-JAN-1995.
PT 15-JUL-1994; U07926.
PR 15-JUL-1993; US-091941.
PS 26-JUL-1993; US-097354.
PA (INCY-) INCYTE PHARM INC.
DR (INDV) UNIV INDIANA FOUND.
PI Seilhamer JJ, Tischfield JA;
DR WPI; 95-067096/09.
N-PSDB: 081139.
PT Novel type III and IV low mol. wt. phospholipase A2 enzymes -
PT from humans and rats, also nucleic acid sequences useful, e.g.
PT for recombinant prodn. of enzymes, research into Batten's
PT disease, etc.
PS Disclosure; Page 67-76; 160pp; English.
CC A human PLA2-encoding cDNA (081136) expressing HPLA2-10, was
CC isolated from human brain RNA by RACE-PCR. 2 Rat PLA2 cDNAs,
CC designated RPLA2-8 (081136) and RPLA2-10 (081137), were isolated from
CC rat brain and heart cDNA libraries, respectively. A partial human
CC genomic counterpart to RPLA2-8, HPLA2-8 (081139), was also obtained.
CC RPLA2-8 and HPLA2-8 have been designated type III PLA2, and RPLA2-10
CC and HPLA2-10 as type IV.
SQ Sequence 146 AA;

QY 13 CSPKOTEGHRCESCMALLSSS 36

RESULT 15
ID R1108 standard; Protein; 555 AA.
AC R1108;
DE 23-MAY-1991 (first entry)
DE Luciferase deduced from cDNA.
KW Vargula hilgendorffii.
OS Vargula hilgendorffii.
PN J03030678-A.
PD 08-FEB-1991.
PF 29-JUN-1989; 167689.
PR 29-JUN-1989; JP-167689.
PA (OHSA-) OHSKA BIOSCIENCE K.
DR WPI; 91-084343/12.
DR N-PSDB; 010957.
PT DNA cfd. coding luciferase derived from Vargula hilgendorffii - by
transfection of host cells and culturing.
PS Claim 1; Fig 3; 11pp; Japanese.
CC The cDNA encoding the protein is used to prepare vector pRSVVL
which is used to transfect COS cells (ATCC CRL1650) for expression
of luciferase. The enzyme is useful for assays in biomedical or
environmental fields.
SQ Sequence 555 AA;

Query Match 18.3%; Score 69; DB 2; Length 555;
Best Local Similarity 33.3%; Pred. No. 2.26e+01;
Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 39 ceakegecidrcatckridlsdg 62
QY 13 CSPKOTEGHRCESCMALLSSS 36

Search completed: Tue Mar 19 09:55:19 1996
Job time : 9 secs.

W P E S R E H
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Mar 19 09:54:27 1996; MasPar time 3.63 Seconds
Tabular output not generated. 341.191 Million cell updates/sec

Title: >US-08-404-832-3
Description: (1-49) from US08404832.pep
Perfect Score: 377
Sequence: 1 DKYKCEKCHLVICSPKOTEGHRCESCMALLSSSPKTCACESIVK 49

Scoring table: PAM 150
Gap 11

Searched: 82306 seqs, 25270970 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p146
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4
8:unann5 9:unann6 10:unann7 11:unann8 12:unrev1
13:unrev2

Statistics: Mean 30.545; Variance 66.324; scale 0.461

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	377	100.0	568	11	A55649	TNFR-associated prot	8.71e-52
2	120	31.8	377	11	A47380	KING finger-contain	8.43e-07
3	109	28.9	1863	10	A54652	breast/ovarian cance	3.93e-05
4	107	28.4	475	2	A37241	52K autoantigen Ro/S	7.80e-05
5	105	27.9	1042	11	S4379	RAG-1 protein - rabb	1.54e-04
6	105	27.9	1042	11	S42511	RAG-1 protein - rabb	1.54e-04
7	102	27.1	259	9	S53400	18543.3 protein - ye	4.24e-04
8	102	27.1	487	3	DDBY18	DNA repair protein R	4.24e-04
9	102	27.1	1043	11	A33754	recombination-activa	4.24e-04
10	99	26.3	501	12	S34825	uvr2 protein - Neuro	1.16e-03
11	99	26.3	609	10	A43906	nuclear phosphoprote	1.16e-03
12	98	26.0	1043	11	B33754	recombination-activa	1.16e-03
13	97	25.7	477	9	S55494	C3HC4 type zinc fing	2.24e-03
14	94	24.9	328	9	S00530	surface antigen CRP1	5.99e-03
15	94	24.9	498	12	B42125	cysteine-rich surfac	5.99e-03
16	94	24.9	624	10	S28418	probable zinc-bindin	5.99e-03
17	94	24.9	1009	10	A56095	H1p16 protein - hum	5.99e-03
18	94	24.9	1009	10	S49618	helicase-like transc	5.99e-03
19	94	24.9	1041	13	S42509	Rag-1 protein - chic	5.99e-03
20	90	23.9	160	10	S28290	hypothetical protein	2.18e-02
21	85	22.5	658	8	A44272	photomorphogenesis r	1.06e-01
22	85	22.5	834	10	S44866	ROS53.4 protein - Ca	1.06e-01
23	85	22.5	1169	9	S31301	DNA repair protein R	1.06e-01
24	83	22.0	315	6	D48560	immediate-early prot	1.97e-01
25	83	22.0	398	7	S34203	heat shock protein 7	1.97e-01
26	83	22.0	598	7	S28712	heat shock protein 7	1.97e-01
27	83	22.0	630	10	A49656	estrogen-responsive	1.97e-01
28	83	22.0	775	3	EDBE11	immediate-early prot	1.97e-01
29	82	21.8	361	13	S17880	gene posterior sex c	2.68e-01
30	82	21.8	365	11	A30891	regulatory protein r	2.68e-01
31	82	21.8	1603	10	S17963	gene posterior sex c	2.68e-01

```
32 81 21.5 713 9 A35502 major surface-labele 3.64e-01
33 79 21.0 208 13 S38745 cytochrome-rich protei 6.68e-01
34 79 21.0 670 9 S53414 L9576.3 protein - ye 6.68e-01
35 78 20.7 141 9 PC1294 trophozoite surface 9.02e-01
36 78 20.7 798 6 B28193 integrin beta-1* cha 9.02e-01
37 78 20.7 798 6 A28193 integrin beta-1 chal 9.02e-01
38 78 20.7 1548 13 S34583 serine proteinase (E 1.22e+00
39 77 20.4 324 11 A39523 transforming protein 1.22e+00
40 76 20.2 560 11 A40044 PML-1 protein - huma 1.64e+00
41 76 20.2 659 1 TVMSBF protein kinase (EC 2 1.64e+00
42 76 20.2 802 11 S44382 PML-3 protein - huma 1.64e+00
43 76 20.2 802 11 S42518 PML-3 protein - huma 1.64e+00
44 76 20.2 860 11 S44380 PML-1 protein - huma 1.64e+00
45 76 20.2 1429 10 S06434 homeotic protein 1ln 1.64e+00
```

ALIGNMENTS

```
RESULT 1
ENTRY A55649 #type complete
TITLE TNFR-associated protein LBP1 - human
ALTERNATE_NAMES CD40-binding protein
ORGANISM #formal name Homo sapiens #common name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 03-Apr-1995
```

```
ACCESSIONS
REFERENCE A55649; A55135
#authors Mosialos, G.; Birkenbach, M.; Valamanchi, R.; Vamvakale, T.; Wate, C.; Kieff, E.
```

```
#journal Cell (1995) 80:389-399
#title The Epstein-Barr virus transforming protein LBP1 engages signaling proteins for the tumor necrosis factor receptor family.
```

```
#accession A55649
#molecule_type mRNA
#residues 1-568 #label MOS
#cross-references GB:019260
#note nucleotide sequence not given
```

```
REFERENCE A55135
#authors Hu, H.M.; O'Rourke, K.; Bogutski, M.S.; Dixit, V.M.
#journal J. Biol. Chem. (1994) 269:30069-30072
#title A novel RING finger protein interacts with the cytoplasmic domain of CD40.
```

```
#accession A55135
#molecule_type mRNA
#residues 1-133 135-404, 'C', 406-568 #label HDA
#cross-references GB:015637
#note nucleotide sequence not given
```

```
KEYWORDS coiled coil
```

```
SUMMARY #region RING-finger motif
#length 568 #molecular-weight 64490 #checksum 8660
```

```
Query Match 100.0% Score 377, DB 11; Length 568;
Best Local Similarity 100.0% Pred. No. 8,71e-52;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 49 dkycckchlvicspkctecrfccscmaallssapctacqesivk 97
1 DKYCKCHLVICSPKCTECRFCCSCMAALLSSSPKCTACQESIVK 49
```

```
RESULT 2
ENTRY A47380 #type complete
TITLE RING finger-containing DNA binding protein RING1 - human
ALTERNATE_NAMES #formal name Homo sapiens #common name man
ORGANISM #formal name Homo sapiens #sequence_revision 18-Nov-1994 #text_change 21-Jan-1994
```

```
ACCESSIONS
REFERENCE A47380
#authors Lowering, R.; Hanson, I.M.; Borden, K.L.; Martin, S.; O'Reilly, N.J.; Evans, G.I.; Rahman, D.; Pappin, D.J.; Trowdale, J.; Freeman, P.S.
```

```
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:2112-2116
#title Identification and preliminary characterization of a protein motif related to the zinc finger.
```

```
#cross-references MIMD:93211912
```

```
#accession A47380
#status preliminary
#molecule_type DNA; protein
#residues 1-377 #label LOV
#cross-references NCBI:128010; NCBI:P128011
```

```
SUMMARY #note sequence extracted from NCBI backbone
#length 377 #molecular-weight 39145 #checksum 7895
```

```
Query Match 31.8% Score 120; DB 11; Length 377;
Best Local Similarity 34.8% Pred. No. 8,45e-07;
Matches 16; Conservative 11; Mismatches 18; Indels 1; Gaps 1;
```

```
Db 18 mcpielmkmttkeclhrfscdeivrlsgnkepcrtkklv 63
4 KCEKCHLVICSPKCTECRFCCSCMAALLSSSPKCTACQESIV 48
```

```
RESULT 3
ENTRY A54652 #type complete
TITLE breast/ovarian cancer susceptibility protein BRCA1 - human
ORGANISM #formal name Homo sapiens #common name man
DATE 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 08-Dec-1994
```

```
ACCESSIONS A54652
REFERENCE A54652
#authors Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.;
```

```
Hartshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L.M.; Ding, W.; Bell, R.; Rosenthal, J.; Husey, C.; Tran, T.; McClure, M.; Frye, C.; Hattier, T.; Phelps, R.; Haugen-Strano, A.; Katcher, H.; Yakumo, K.; Gholami, Z.; Shaffer, D.; Stone, S.; Bayer, S.; Wray, C.; Bogden, R.; Dayananth, P.; Ward, J.; Tonlin, P.; Nared, S.; Bristow, P.K.; Norris, F.H.; Helvering, L.; Morrison, P.; Roelcke, P.; Lai, M.; Barrett, J.C.; Lewis, C.; Neuhausen, S.; Cannon-Albright, L.; Goldgar, D.; Wiseman, R.; Kamb, A.; Skolnick, M.H.
Science (1994) 266:66-71
A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.
```

```
#accession A54652
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-1863 #label M1K
```

```

GENETICS      ##cross-references GB:014680
#gene         GDB:BRCA1
#map_position 17q12-q21
KEYWORDS      zinc finger
SUMMARY       #length 1863 #molecular-weight 207755 #checksum 1680

Query Match      28.9%; Score 109; DB 10; Length 1863;
Best Local Similarity 39.6%; Pred. No. 3.93e-05;
Matches 19; Conservative 3; Mismatches 22; Indels 4; Gaps 4;

Db      24 cpic-1elikepvatckdhifckfcmklmgkqkgsqgplckndtck 70
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      5 CEKHIVLCS-PRKTECGHRCFCSGMAALLS-SSSP-KCTACQGEIVK 49

RESULT 4
ENTRY      A37241 #type complete
TITLE      52K autoantigen Ro/SS-A - human
ALTERNATE_NAMES  Sjogren syndrome antigen A
ORGANISM   Homo sapiens #common name man
DATE       07-Feb-1992 #sequence_revision 26-May-1995 #text_change
30-Jun-1995

ACCESSIONS  A55642; A37241; A37240
REFERENCE   A55642
#authors    Tangu, H.; Horowitz, R.; Gibson, N.; Frank, M.B.
#journal     Genomics (1994) 24:541-548
#title      The location of a disease-associated polymorphism and genomic
            structure of the human 52-kDa Ro/SSA locus (SSA1).
            A55642
            not compared with conceptual translation

#accession  #molecule_type DNA
#residues   1-475 #label TSU
#cross-references GB:013657

REFERENCE   A37241
#authors    Itoh, K.; Itoh, Y.; Frank, M.B.
#journal     J. Clin. Invest. (1991) 87:177-186
#title      Protein heterogeneity in the human Ro/SSA ribonucleoproteins.
            The 52- and 60-kD Ro/SSA autoantigens are encoded by
            separate genes.

#cross-references M01D:91086445
#accession  A37241
#molecule_type mRNA
#residues   1-475 #label ITO
#cross-references GB:M34551

REFERENCE   A37240
#authors    Chan, E.K.-L.; Hamel, J.C.; Buyn, J.P.; Tan, E.M.
#journal     J. Clin. Invest. (1991) 87:68-76
#title      Molecular definition and sequence motifs of the 52-kD
            component of human SS-A/Ro autoantigen.

#cross-references M01D:91086480
#accession  A37240
#molecule_type mRNA
#residues   1-51, 'A', '53-475 #label CHA
#cross-references GB:M62800; GB:M35041

GENETICS      GDB:SSA1
#gene         #map_position 11p15.5-p15.3
#introns      136/3, 168/3, 245/3, 253/2, 287/1
CLASSIFICATION #superfamily rfp transforming protein

```

KEYWORDS	DNA binding: nucleolus; zinc finger
SUMMARY	#length 475 #score 107; DB 2; Length 475; #checksum 7482
Query Match	28.4%; Score 107; DB 2; Length 475;
Best Local Similarity	27.3%; Pred. No. 7,80e-05;
Matches 12; Conservative 13; Mismatches 19; Indels 0; Gaps 0;	
Db	16 cpicldpfvvpvaiecgahfcgdciaisygkvqgsvcpvcrfll 59
Qy	5 CERCHVLCSPKQTECGHRCESCRNALSSSPKCAQESIV 48
RESULT	5
ENTRY	S44379 #type complete
TITLE	RAC-1 protein - rabbit
ORGANISM	#formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE	06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
ACCESSIONS	S44379
REFERENCE	S44379
#authors	Fuschiotti, P.; Harindranath, N.; Mage, R.G.; McCormack, W.T.; Dhanarajan, P.; Roux, K.H.
#submission	submitted to the EMBL Data Library, December 1991
#description	Recombination activating genes -1 and -2 of the rabbit: cloning and characterization of germ-line and expressed genes.
#accession	S44379
#status	preliminary
#molecule_type	DNA
#residues	1-1042 #label FUS
#cross-references	EMBL:J77666
SUMMARY	#length 1042 #molecular_weight 119006 #checksum 8825
Query Match	27.9%; Score 105; DB 11; Length 1042;
Best Local Similarity	35.0%; Pred. No. 1.54e-04;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;	
Db	292 cqlcehladdvpetcshvfricilrcilkykmgsyvcpsq 331
Qy	5 CERCHVLCSPKQTECGHRCESCRNALSSSPKCAQ 44
RESULT	6
ENTRY	S42511 #type complete
TITLE	RAC-1 protein - rabbit
ORGANISM	#formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE	13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Jan-1995
ACCESSIONS	S42511
REFERENCE	S42511
#authors	Fuschiotti, P.; Harindranath, N.; Mage, R.G.; McCormack, W.T.; Dhanarajan, P.; Roux, K.H.
#journal	Mol. Immunol. (1993) 30:1021-1032
#title	Recombination activating genes-1 and -2 of the rabbit: cloning and characterization of germ-line and expressed genes.
#accession	S42511
#status	preliminary

[illegible]

```

REFERENCE      S22260
#authors      Benit, P.; Chanet, R.; Fabre, F.; Faye, G.; Fukuhara, H.;
#journal      Sor, F.
#title        Yeast (1992) 8:147-153
#cross-references      Sequence of the sup61-RAD18 region on chromosome III of
                        Saccharomyces cerevisiae.
#accession    S22263
##molecule_type      DNA
##residues      1-487 ##label      BEN
#cross-references      EMBL:S59398
REFERENCE      S19477
#authors      Antoku, G.; Benit, P.; Chanet, R.; Fabre, R.; Faye, G.;
                        Fukuhara, H.; Mathieu, A.; Sor, F.
#submission    submitted to the Protein Sequence Database, March 1992
#accession    S19481
##molecule_type      DNA
##residues      1-487 ##label      ANT
#cross-references      EMBL:X59720
REFERENCE      JS0082
#authors      Chanet, R.; Magana-Schwencke, N.; Fabre, F.
#journal      Gene (1988) 74:343-347
#title        Potential DNA-binding domains in the RAD18 gene product of
                        Saccharomyces cerevisiae.
#cross-references      M01D:89232745
#accession    JS0082
##molecule_type      DNA
##residues      1-487 ##label      CHA

GENETICS
#gene          LISTA:RAD18
#map_position  3R
CLASSIFICATION #superfamily DNA repair protein RAD18
KEYWORDS       ATP; DNA binding; DNA repair; zinc finger
FEATURE
28-48          #region zinc finger motif\
51-65          #region zinc finger motif\
190-210        #region zinc finger motif\
366            #binding site ATP (Lys) #status predicted

SUMMARY
#length 487 #molecular-weight 55230 #checksum 1135

Query Match      27.1%; Score 102; DB 3; Length 487;
Best Local Similarity 37.5%; Pred. No. 4.24e-04;
Matches 15; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

Db      27 rchckdflkypvltpcgtfscloirhng-pncplc 65
      :| | | | | | | | | | | | | | | | | | |
Oy      4 KCEKCHIVLCSPKQITCGHRCESMALLSSSPKCTAC 43

RESULT      9
ENTRY      A33754 #type complete
TITLE      recombination-activating protein 1 - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
                        31-Dec-1993
A33754
A33754      Schatz, D.G.; Oettinger, M.A.; Baltimore, D.
REFERENCE   Cell (1989) 59:1035-1046
#journal    The V(D)J recombination activating gene, RAG-1.
#authors
#title

```



```

# #cross-references NCBI:n64515; NCBI:P:64520; EMBL:R63705
# #note sequence extracted from NCBI backbone
GENETICS
# #gene knf7
KEYWORDS DNA binding; phosphoprotein; zinc finger
SUMMARY length 609 #molecular-weight 68875 #checksum 6416

Query Match 26.3%; Score 99; DB 10; Length 609;
Best Local Similarity 37.5%; Pred. No. 1.16e-03;
Matches 18; Conservative 6; Mismatches 23; Indels 1; Gaps 1;

Db 141 eeltcplcvelfkdpvmacgnfciscidkawsqgsfaepecteei 188
: | | :: | | | | | | | | | | | | | | | | | | |
Oy 1 DKRCCKHLVLCSPKOTCGCHRCSCMA-ALLSSSPKTCACQESI 47

RESULT 12
ENTRY B33754 #type complete
TITLE recombination-activating protein 1 - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
31-Dec-1993
ACCESSIONS B33754
REFERENCE A33754
#authors Schatz, D.G.; Oettinger, M.A.; Baltimore, D.
#journal Cell (1989) 59:1035-1048
#title The V(D)J recombination activating gene, RAG-1.
#cross-references MIMD:90090604
#accession B33754
# #molecule_type mRNA
# #residues 1-1043 #label SCA
KEYWORDS #cross-references GB:M29475
SUMMARY DNA binding; zinc finger
#length 1043 #molecular-weight 119490 #checksum 914

Query Match 26.0%; Score 98; DB 11; Length 1043;
Best Local Similarity 32.5%; Pred. No. 1.61e-03;
Matches 13; Conservative 8; Mismatches 19; Indels 0; Gaps 0;

Db 293 cgcicelladpyetckhlfrcilrcrlkmgyspcer 332
| : | | | | | | | | | | | | | | | | | | |
Oy 5 CERCHLVLCSPKOTCGCHRCSCMAALLSSSPKTCACQ 44

RESULT 13
ENTRY S55494 #type complete
TITLE C3H/c type zinc finger protein - Emericella nidulans
ORGANISM #formal name Emericella nidulans, Aspergillus nidulans
DATE 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change
01-Sep-1995
ACCESSIONS S55494
REFERENCE S55494
#authors Iwamajko, L.A.; Cotton, C.M.; Jones, G.W.; Tomsett, A.B.;
Strike, P.
#submission submitted to the EMBL Data Library, June 1995
#description Cloning and characterization of nuva, an Aspergillus nidulans
#accession S55494 gene involved in DNA repair and recombination.
# #status preliminary
# #molecule_type mRNA

```


and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description	Pred. No
1	176	46.7	501	TRF2_MOUSE	TNF RECEPTOR ASSOCIAT	1.71e-19
2	120	31.8	377	RINI_HUMAN	RING1 PROTEIN.	1.11e-08
3	109	28.9	1863	BRC1_HUMAN	BREAST CANCER TYPE 1	1.06e-06
4	107	28.4	475	ROS2_HUMAN	52 KD RO PROTEIN (SUO	3.38e-06
5	105	27.9	1042	RAG1_RABIT	V(D)J RECOMBINATION A	5.32e-06
6	102	27.1	487	RAG1_YEAST	DNA REPAIR PROTEIN RA	1.76e-05
7	102	27.1	1043	RAG1_HUMAN	V(D)J RECOMBINATION A	1.76e-05
8	99	26.3	501	UVS2_NDCCR	UVS-2 PROTEIN.	5.76e-05
9	98	26.0	1040	RAG1_MOUSE	V(D)J RECOMBINATION A	8.52e-05
10	94	24.9	328	C17O_GIALA	SURFACE ANTIGEN CRP17	4.01e-04
11	94	24.9	624	A33_FLEMA	ZINC-BINDING PROTEIN A	4.01e-04
12	94	24.9	1041	RAG1_CHICK	V(D)J RECOMBINATION A	4.01e-04
13	90	23.9	160	YNC1_CAEEL	HYPOTHETICAL 18.7 KD	1.84e-03
14	85	22.5	834	YNC1_CAEEL	HYPOTHETICAL 97.1 KD	1.84e-02
15	85	22.5	1169	RAD51_YEAST	HEAT SHOCK PROTEIN 70	2.43e-02
16	83	22.0	598	HST1_SERV	HEAT SHOCK PROTEIN 70	2.43e-02
17	83	22.0	775	ICP0_HSV11	TRANS-ACTING TRANSCI	3.49e-02
18	82	21.8	353	RPT1_MOUSE	DOWN REGULATORY PROTE	3.49e-02
19	82	21.8	1603	PSC_DROME	POSTERIOR SEX COMBS P	3.49e-02
20	81	21.5	713	TS4A_GIALA	MAJOR SURFACE-LABELLED	5.00e-02
21	79	21.0	208	CRP2_RAT	CYSSTEINE-RICH PROTEIN	1.02e-01
22	79	21.0	555	DP87_DICD1	DP87 PROTEIN.	1.02e-01
23	78	20.7	798	ITB1_XENLA	INTEGRIN BETA-1 SUBUN	1.44e-01
24	78	20.7	798	ITB0_XENLA	INTEGRIN BETA-1* SHED	1.44e-01
25	77	20.4	324	EM11_MOUSE	DNA-BINDING PROTEIN B	2.05e-01
26	77	20.4	326	EM11_HUMAN	DNA-BINDING PROTEIN B	2.05e-01
27	76	20.2	274	YE06_YEAST	HYPOTHETICAL 30.8 KD	2.89e-01
28	76	20.2	560	PM1B_HUMAN	PUTATIVE TRANSCRIPTO	2.89e-01
29	76	20.2	589	Y041_CAEEL	HYPOTHETICAL 68.7 KD	2.89e-01
30	76	20.2	611	PM1L_HUMAN	PROBABLE TRANSCRIPTO	2.89e-01
31	76	20.2	633	PM1X_HUMAN	PROBABLE TRANSCRIPTO	2.89e-01
32	76	20.2	824	PM1L_HUMAN	PROBABLE TRANSCRIPTO	2.89e-01
33	76	20.2	882	PM1L_HUMAN	PROBABLE TRANSCRIPTO	2.89e-01
34	76	20.2	1429	L112_CAEEL	LIN-12 PROTEIN PRECOR	2.89e-01
35	75	19.9	115	YBR2_YEAST	HYPOTHETICAL 13.2 KD	4.08e-01
36	75	19.9	513	PSV_DICD1	PRESPORE VESICLE PROT	4.08e-01
37	74	19.6	803	ITB1_CHICK	INTEGRIN BETA-1 PRECU	5.73e-01
38	74	19.6	1801	LMR2_RAT	LAMRIN BETA-2 CHAIN	5.73e-01
39	73	19.4	138	PAZB_VIPAA	PHOSPHOLIPASE A2 PREC	8.03e-01
40	73	19.4	138	PAZC_VIPAA	PHOSPHOLIPASE A2 PREC	8.03e-01
41	73	19.4	138	PAZA_VIPAA	PHOSPHOLIPASE A2 PREC	8.03e-01
42	73	19.4	236	GLM1_TYLA	GIANT EXTRACELLULAR H	8.03e-01
43	73	19.4	342	ME18_MOUSE	DNA-BINDING PROTEIN M	8.03e-01
44	73	19.4	344	ME18_HUMAN	DNA-BINDING PROTEIN M	8.03e-01
45	73	19.4	410	ICP0_PNVIF	TRANS-ACTING TRANSCI	8.03e-01

ALIGNMENTS

RESULT 1			
ID	TRF2_MOUSE	STANDARD;	PRT; 501 AA
AC	P39429;		
DT	01-FEB-1995 (REL. 31, CREATED)		

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE TNF RECEPTOR ASSOCIATED FACTOR 2 (TRAF2).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA;
CC EUTHERIA; RODENTIA;
CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

RP SEQUENCE FROM N.A.

RA ROTHE M., WONG S.C., HENZEL W.J., GOEDEL D.V.,
 DT CERT 78-691 603/100A)

CC -!- FUNCTION: SIGNAL TRANSDUCER ASSOCIATED WITH THE CYTOPLASMIC DOMAIN
CC OF THE 75 KD TUMOR NECROSIS FACTOR RECEPTOR (TNF-R2).
CC

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- SUBUNIT: HELIXDIMER OF IRA1 AND IRA2.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER

DR EMBL; L35303; MMTRAF2A.

KW ZINC-FINGER; COILED COIL.

FT	ZN FING	34	72	C3HC4-TYPE
100	100	100	100	100

SEQUENCE	301 AA;	36026 MW;	1264823 CN;
SD			

Query Match	46.7%;	Score 176;	DB 7;	length 501;
Best Local Similarity	45.2%;	Pred. No. 1.71e-19;		
Matches	19;	Conservative	12;	Mismatches 11;
			Indels	0;
			Gaps	0;

Db 31 kylsacknllrrpfaqcqhryscfltsllssgpncaac 72

0y 2 KYKCEKHLVLCSPKQTECGHRCESCMALLSSSSPKCTAC 43

RESULT	2	
ID	RIN1 HUMAN	STANDARD; PRT; 377 AA.

AC Q06587; 01 THU 1004 /DET 30 CREATED)

DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DI	01-JUN-1994 (REL. 29, CREATED)

DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)

DE RING1 PROTEIN.

GN RING1.

US HOMO SAPIENS (HUMAN).
OC ETRAPYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RM	93211912	
DA	10/07/2013	11:00

BA EVAN G I BAHMAN D PAPPIN D J C TROWSDALE J FREEMONT P S :
KA LOVERING K, HANSON I M, BORDEN A L B, MAKIN S, O'NEILL N O :

PROC. NATL. ACAD. SCI. U.S.A. 90:2112-2116(1993).

CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -!- SIMILARITY: CONTAIN

DR EMBL; Z14000; HSRLNG1
 DR PDB; A47390; A47390

DB HSSP: P28990: 1CHC.
DK FIK, A4/380, A4/380.

PROSITE: PS00518; ZINC FINGER C3HC4

KW ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.

FT	ZN FING	FT	58	C3HC4-TYPE.
175	031	031		031

ET	285	GLY-RICH.
ET	231	GLY-RICH.
DOMAIN	348	GLY-RICH.
DOMAIN	1/6	GLY-RICH.
DOMAIN	285	GLY-RICH.

FT	172	175	NUCLEAR LOC
DOMAIN			
283	250		ORI ARROW

SQ SEQUENCE 377 AA; 39145 MW; 716288 CN;

[illegible][illegible]

QY 5 CEKCHLVLCSPKOTEGHRCFCSMAALLSSSPKCTACQ 44

RESULT 10
ID C170 G1A1A STANDARD; PRT; 328 AA.
AC P15739;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DE SURFACE ANTIGEN CRP170 (FRAGMENT).
OS GIARDIA LAMBLIA.
OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; DIPLOMONADIDA;
OC HEXAMITIDE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE WB;
RM 88089405
RA ADAM R.D., AGGARWAL A., TAL A.A., DE LA CRUZ V.F., MCCUTCHAN T.,
RA NASH T.E.;
RL J. EXP. MED. 167:109-118(1988).
CC -1- CSTEINE-RICH, ANTIGENICALLY VARIANT SURFACE PROTEIN.
DR EMBL; X06741; GICRP170.
DR PIR; S00530; S00530.
DR HSSP; P04355; AMT2.
KM REPEAT; ANTIGEN.
FT NON TER 1
FT REPEAT 38 102
FT REPEAT 103 167
FT NON TER 328 328
SQ SEQUENCE 328 AA; 33438 MW; 499276 CN;

Query Match 24.9%; Score 94; DB 1; Length 328;
Best Local Similarity 29.4%; Pred. No. 4.01e-04;
Matches 10; Conservative 10; Mismatches 13; Indels 1; Gaps 1;

Db 211 cpaatpgc-bascdgctenmtgacdcctckeg 243
QY 13 CSPKOTEGHRCFCSMAALLSSSPKCTACQES 46

RESULT 11
ID A33 PLEMA STANDARD; PRT; 624 AA.
AC 002084;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ZINC-BINDING PROTEIN A33.
OS PLEURODELES WALTII (IBERIAN RIBBED NEWT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA;
OC CAUDATA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RM 93154311
RA BELLINI M., LACROIX J.-C., GALL J.G.;
RL EMBL J. 12:107-114(1993).
CC -1- FUNCTION: MAY BE A NUCLEAR REGULATORY PROTEIN THAT IS STORED
CC IN THE GERMINAL VESICLE FOR USE DURING EARLY EMBRYOGENESIS
CC AND MAY PLAY A ROLE IN THE SYNTHESIS OR PROCESSING OF PRE-MRNA
CC DURING OOGENESIS.

CC -1- DEVELOPMENTAL STAGE: IT FIRST APPEARS ON THE CHROMOSOME LOOPS
CC AND IN THE NUCLEOLAR PLASMA OF THE GERMINAL VESICLE (GV). IT IS
CC TRANSMITTED TO THE EGG AT GV BREAKDOWN AND APPEARS IN EMBRYONIC
CC NUCLEI AT THE MID-BLASTULA STAGE AND IS FOUND IN MANY BUT NOT
CC ALL NUCLEI AT STILL LATER STAGES OF EMBRYOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; U04190; PWPWA33A.
KW ZINC-FINGER; NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; COILED COIL;
KM HEPAD REPEAT PATTERN; RNA-BINDING.
FT DOMAIN 19 26
FT DOMAIN 134 149 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN FING 162 201 C3HC4-TYPE.
FT DOMAIN 238 269 B BOX.
FT DOMAIN 337 386 COILED COIL.
FT DOMAIN 449 619 RFP-LIKE.
SQ SEQUENCE 624 AA; 71056 MW; 2023970 CN;

Query Match 24.9%; Score 94; DB 1; Length 624;
Best Local Similarity 32.6%; Pred. No. 4.01e-04;
Matches 15; Conservative 7; Mismatches 23; Indels 1; Gaps 1;

Db 158 edltcplrtsfkcpvllccgmthckhckidskwnesatgpcck 203
QY 1 DKYKCEKCHLVLCSPKOTEGHRCFCSMAALLSSSP-KCTACQ 45

RESULT 12
ID RAG1 CHICK STANDARD; PRT; 1041 AA.
AC P24271;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE V(D)J RECOMBINATION ACTIVATING PROTEIN.
GN RAG-1.
OS GALUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RM 91098648
RA CARLSON L.M., OETTINGER M.A., SCHATZ D.G., MASTELLER E.L.,
RA HURLEY E.A., MCCORMACK W.I., BALTIMORE D., THOMPSON C.B.;
RL CELL 64:201-208(1991).
CC -1- FUNCTION: RAG1 & RAG2 SYNERGISTICALLY ACTIVATE THE IMMUNOGLOBULIN
CC V-D-J RECOMBINATION. V-D-J RECOMBINATION IS THE COMBINATORIAL
CC PROCESS BY WHICH DEVELOPING LYMPHOCYTES BEGIN TO GENERATE THEIR
CC ENORMOUS RANGE OF BINDING SPECIFICITIES FROM A LIMITED AMOUNT OF
CC GENETIC INFORMATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; M58530; GGRAG1G.
DR PIR; S42509; S42509.
DR PROSITE; P500518; ZINC FINGER C3HC4.
KW ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
FT ZN FING 284 327 C3HC4-TYPE.
SQ SEQUENCE 1041 AA; 119916 MW; 5442207 CN;

Query Match 24.9%; Score 94; DB 6; Length 1041;
Best Local Similarity 30.8%; Pred. No. 4.01e-04;

CC		Maintenance of wild-type rates of instability of simple repetitive sequences such as poly(GT) repeats.
CC	-1-	SUBCELLULAR LOCATION: NUCLEAR.
CC	-1-	SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR	EMBL; M96644;	SCRAD5A.
DR	EMBL; S43248;	S43248.
DR	EMBL; S46103;	S46103.
DR	PIR; S31301;	S31301.
DR	PIR; S26983;	S26983.
DR	LISTA; SC009241;	RAD5.
DR	PROSITE; PS00538;	ZINC_FINGER_C3HC4.
KW	DNA DAMAGE; DNA REPAIR; NUCLEAR PROTEIN; ZINC-FINGER; DNA-BINDING; HELICASE; ATP-BINDING.	
FT	DOMAIN	42..60 ASP/GLD-RICH (ACIDIC).
FT	DOMAIN	303..315 ARG/LYS-RICH (BASIS).
FT	NP BIND	532..539 ATP (POTENTIAL).
FT	SITE	681..684 DEGH BOX.
FT	ZN FING	914..960 C3HC4-TYPE.
FT	CONFLICT	478..478 O -> R (IN REF. 2).
FT	CONFLICT	635..635 T -> N (IN REF. 2).
FT	CONFLICT	846..846 G -> S (IN REF. 2).
FT	CONFLICT	898..898 R -> S (IN REF. 2).
FT	CONFLICT	973..973 V -> A (IN REF. 2).
FT	CONFLICT	1063..1063 A -> R (IN REF. 2).
RSQ	SEQUENCE	1169 AA; 134001 MW; 7092987 CN;
Db	930 techsfcekccl	941
Qy	18 TECHRFCESCM	29
	:	
	Query Match	22.5%; Score 85; DB 6; Length 1169;
	Beat Local Similarity	75.0%; Pred. No. 1,19e-02;
	Matches 9; Conservative	1; Mismatches 2; Indels 0; Gaps 0;

Search completed: Tue Mar 19 09:54:09 1996
Job time : 7 secs.